WEIGHTED INEQUALITIES VIA DYADIC OPERATORS AND A LEARNING THEORY APPROACH TO COMPRESSIVE SENSING

A Dissertation
Presented to
The Academic Faculty

By

Timothy Scott Spencer

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
School of Mathematics

Georgia Institute of Technology

August 2017

WEIGHTED INEQUALITIES VIA DYADIC OPERATORS AND A LEARNING THEORY APPROACH TO COMPRESSIVE SENSING

Approved by:

Dr. Michael Lacey, Advisor
School of Mathematics
Georgia Institute of Technology

Dr. Shahaf Nitzan
School of Mathematics
Georgia Institute of Technology

Dr. Galyna Livshyts
School of Mathematics
Georgia Institute of Technology

Dr. Doron Lubinsky
School of Mathematics
Georgia Institute of Technology

Dr. Mark A. Davenport
School of Electrical and Computer Engineering
Georgia Institute of Technology

Date Approved: March 30, 2017

"Politicians use statistics in the same way that a drunk uses lamp-posts - for support rather than illumination."

-Andrew Lang

ACKNOWLEDGEMENTS

Before any acknowledgements, I apologize to the deserving many that I will negligently omit. I thank first my advisor, Michael Lacey; without his guidance, support and encouragement I would've been lost throughout my graduate tenure. He never failed to share engaging and fruitful research problems, interesting perspectives, and colorful insights to life in general. His contributions as a coauthor were invaluable.

I also thank Shahaf Nitzan for not only serving on my oral exam and thesis defense committees, but for her many encouragements and assurances, for convincing me to give talks and remain mathematically active when I was in need of motivation, and for always tolerating my perpetual tardiness. I also give a special thanks to Brett Wick for his insight and guidance in my early and his late stages at Georgia Tech.

I owe a special thanks to Robert Rahm who was not only a good friend, but an extremely bright and productive coauthor; much of the work here is done in collaboration with him. On the same note, I thank Irina Holmes for her indispensable time, creativity, and commitment as a coauthor.

I also thank the other members of the committee – Galyna Livshyts, Doron Lubinsky and Mark Davenport – whose valuable time was plenty an excuse to refuse my request.

Lastly and probably leastly considering the context, I thank my family and friends for making my life as a graduate student a bit more bearable.

TABLE OF CONTENTS

LIST OF FIGURES

SUMMARY

This thesis explores topics from two distinct fields of mathematics. The first part addresses a theme in abstract harmonic analysis, while the focus of the second part is a topic in compressive sensing.

The first part of this dissertation explores the application of dominating operators in harmonic analysis by *sparse* operators. In the second chapter, we introduce sparse operators. Presented therein are preliminary results on dominating certain operators by sparse operators, and we also prove several analogous results for other operators that we use in later chapters. The results were achieved in collaboration with the coauthors credited for corresponding chapters. The third chapter concerns Calderón-Zygmund operators. We make use of the sparse domination introduced in Chapter 2 to derive weighted inequalities for Calderón-Zygmund operators and the Hardy-Littlewood maximal operator. This chapter comprises results that were established in independent collaborations with Michael Lacey and Robert Rahm. Chapter 4 establishes weighted inequalities for the fractional integral operators (also known as Riesz potentials) and fractional maximal operator. These results were also achieved in collaboration with Robert Rahm. Chapter 5 deviates from the theme of domination by sparse operators, but continues the study of fractional integral operators. There is a another notion from dyadic calculus used here, namely averaging over dyadic operators. We use these methods to achieve weighted inequalities for commutators of fractional integral operators with multiplication operators. An interesting result is that the inequality can be reversed. Since the bound depends on a BMO norm of the function in the multiplication operator, we characterize a certain BMO space by the boundedness of the commutator with fractional integral operators. This work was done in collaboration with Robert Rahm and Irina Holmes. Chapter 6 addresses oscillatory integral

operators and random discrete Hilbert transforms. The oscillatory integrals are built by polynomial modulation of Calderón-Zygmund kernels. For both of these classes of operators, we establish sparse bilinear bounds that induce weighted inequalities. In the case of the random discrete Hilbert transforms, these are believed to be the first results of their kind. This work is done in collaboration with Michael Lacey.

In the second part, we explore the utility of learning theory in the relatively new field of compressive sensing. The focus is on the subfield of one-bit sensing. Chapter 7 briefly introduces the pertinent topics from compressive sensing and demonstrates how a fundamental result of the field can be established using the techniques in Chapter 8. The last chapter contains the point Part II. We introduce the notion of one-bit sensing and an analogue of the Restricted Isometry Property, which is a type of quasi-isometry developed in compressive sensing. We are able to effectively estimate the VC-dimension of hemispheres relative to sparse vectors, which allows us to employ learning theory techniques to control an empirical process. This control implies the desired Restricted Isometry Property with high probability. With these methods, we are also able to discuss the effects of certain noise models on the acquisition scheme.

INTRODUCTION

The first part of this dissertation explores application of dyadic calculus, a fruitful subfield of harmonic analysis. This theme is relatively new, and there are many avenues still to be explored.

In the second chapter, we introduce sparse operators. Preliminary results on dominating certain operators by sparse operators are presented. We also prove several analogous results for other operators that we use in later chapters. The appeal of sparse operators is that they are highly-localized, positive operators, something that the operators they dominate are not. While this comparison is surprising in its own right, the focus is on the application to attain weighted inequalities.

Next, we introduce several related classes of operators: fractional integral operators, Calderón-Zygmund operators, and their related oscillatory and random versions. These are the complicated, non-local operators mentioned above. The sparse bounds allow the easy deduction of weighted inequalities. A weighted inequality is an inequality that bounds the norm of an operator, or a class of operators, by a characteristic of the weights on the spaces the operator maps between. A weight is a non-negative, locally integrable function. Since we treat them as densities of the induced measure, for a weight w we will often write $w(A)$ meaning $\int_A w(x)dx$. There are two techniques of domination that are explored. For a Calderón-Zygmund or fractional integral operator, we dominated the operator pointwise by sparse operators. For an oscillatory singular integral or random Hilbert transform, we dominate the bilinear form $\langle Tf, g\rangle$ by sparse bilinear forms. This is a generalization of the former since the sparse bilinear forms are a generalization of the bilinear forms associated to sparse operators. While the sparse objects in both cases depend on the function function(s) the operator is applied to (see Section 2.2, for instance), weighted inequalities for the *class* of sparse operators extend. In both cases, we are able to deduce new, meaningful weighted inequalities.

Chapter 5 deviates from the theme of domination by sparse operators, but continues the study of fractional integral operators. There is a another notion from dyadic calculus used here, namely averaging over dyadic operators. We use these methods to achieve weighted inequalities for commutators of fractional integral operators with multiplication operators. An interesting result is that the inequality can be reversed. Since the bound depends on a BMO norm of the function in the multiplication operator, we characterize a certain BMO space by the boundedness of the commutator with fractional integral operators.

In the second part, we explore the utility of learning theory in the relatively new field of compressive sensing. The objective of compressive sensing is to exploit low dimensionality properties of certain classes of signals (read high dimensional vectors) to acquire and reconstruct the signals at sub-Nyquist rates. For us, we assume the signals are sparse, i.e. they have relatively few non-zero coordinates. We point out here that this notion of sparseness is unrelated to the one mentioned earlier. The nomenclature is admittedly inconvenient, but it is consistent with the existing literature. We are primarily be interested in extremely quantized measurement, a topic embodied by the subfield one-bit sensing. In this case, only the sign-bit of each measurement is retained. We are able to prove results concerning a quasi-isometry property of the measurement maps by effectively estimating the VC-dimension of the class of hemispheres relative to sparse signals and applying techniques from learning theory.

In both parts, constants are suppressed: by $A \lesssim B$, we mean that there is an absolute, positive constant c so that $A \leqslant cB$. By $A \sim B$, we mean $A \lesssim B$ and $B \lesssim A$.

Part I

Weighted Inequalities via Dyadic Operators

CHAPTER 2

SPARSE DOMINATION

Due to deep and important theorems of Lerner, Lacey, and Rey and Conde–Alonso [56, 47, 16] important operators in harmonic analysis (for example, maximal functions, Calderón–Zygmund Operators, Haar shifts) are pointwise dominated by finite sums of sparse operators. Thus, proving two–weight inequalities for these sparse operators will imply the same theorems for other operators of interest. We begin with some preliminary definitions.

Definition 2.1. A collection $\mathcal{D}$ of cubes in $\mathbb{R}^n$ is said to be a *dyadic grid* if:

(i) Each $Q \in \mathcal{D}$ has side length of 2^k for some $k \in \mathbb{Z}$.

(ii) For $Q, R \in \mathcal{D}$: $Q \cap R$ is measure zero, $Q \subset R$, or $R \subset Q$.

(iii) If $\mathcal{D}_k = \{Q \in \mathcal{D} : \text{the side length of } Q \text{ equals } 2^k\}$, then $\mathbb{R}^n = \cup_{Q \in \mathcal{D}_k} Q$.

Definition 2.2. A subset $\mathcal{S}$ of a dyadic grid is said to be *sparse* if for every $P \in \mathcal{S}$:

$$\sum_{\substack{Q \in \mathcal{S} : Q \subsetneq P \\ Q \text{ is maximal}}} |Q| \leqslant \frac{1}{2} |P| .$$

The portion $\frac{1}{2}$ is arbitrary, and any positive constant less than 1 would work equivalently. That is also true for the following equivalent notion of a sparse collection of cubes which is sometimes more convenient. A collection $\mathcal{S}$ of cubes is sparse if there is a set $E_Q \subset Q$ for each $Q \in \mathcal{S}$ so that

(a) $|E_Q| > \frac{1}{2}|Q|$ for each $Q \in \mathcal{S}$, and

(b) the collection of sets $\{E_Q : Q \in \mathcal{S}\}$ are pairwise disjoint.

Here and throughout, we denote by $\langle f\rangle_Q^\mu$ the μ-average of f on Q: $\mu(Q)^{-1}\int_Q f$. When μ is Lebesgue measure, we simply write $\langle f\rangle$.

Definition 2.3. An operator S is *sparse* if there is a sparse collection of cubes $\mathcal{S}$ so that

$$Sf = \sum_{Q\in\mathcal{S}} \langle f\rangle_Q \mathbf{1}_Q.$$

We typically suppress the dependence of S on the sparse collection $\mathcal{S}$. By abuse of notation, if an operator is sparse with respect to a choice of grid, we call it sparse.

The following deep and useful theorem is due to Sawyer [84]. This result is an integral part of several of the proofs in the following chapters.

Theorem A. *Let $\mathcal{D}$ be a dyadic grid and let $\mathcal{S} \subset \mathcal{D}$ be sparse. Define:*

$$\mathcal{T}_1 := \sup_{P\in\mathcal{S}} \frac{1}{\sigma(P)} \int_P \left| \sum_{Q\in\mathcal{S}:Q\subset P} \langle\sigma\rangle_Q 1_Q(x)\right|^p w(x)dx$$

$$\mathcal{T}_2 := \sup_{P\in\mathcal{S}} \frac{1}{w(P)} \int_P \left| \sum_{Q\in\mathcal{S}:Q\subset P} \langle w\rangle_Q 1_Q(x)\right|^{p'} \sigma(x)dx.$$

Then:

$$\|T_{\mathcal{S}}\sigma\cdot : L^p(\sigma) \to L^p(w)\| \lesssim \mathcal{T}_1^{\frac{1}{p}} + \mathcal{T}_2^{\frac{1}{p'}}.$$

2.1 *Calderón-Zygmund Operators*

A sparse operator is bounded on all L^p, and in fact, is a '*positive* dyadic Calderón-Zygmund operator.' And the class is sufficiently rich to capture the norm behavior of an arbitrary Calderón-Zygmund operator. We use the recent inequality [47], which gives *pointwise control* of a Calderón-Zygmund operator by a sparse operator.

Theorem B. *[47, Thm 4.2] Let T be a Calderón-Zygmund operator and $f \in L^1$ be compactly supported. Then there are at most $N \leqslant 3^d$ sparse operator $S_1, \dots, S_N$ (associated to distinct choices of grids) so that $|Tf| \lesssim \sum_{n=1}^{N} S|f|$.*

As a consequence, it suffices to prove our main theorems on Calderón-Zygmund operators for sparse operators.

2.2 Fractional Integral Operators

In this section, we list several known results related to dominating fractional maximal and fractional integral operators by sparse-like operators; we include some proofs because we could not find them in the literature. Much of this section is taken from [77].

For a given dyadic grid, $\mathcal{D}$, define the dyadic fractional maximal operator:

$$M_\alpha^{\mathcal{D}} f(x) := \sup_{Q \in \mathcal{D}} 1_Q(x) \, |Q|^{\alpha/n} \langle f \rangle_Q$$

and the dyadic fractional integral operator:

$$I_\alpha^{\mathcal{D}} f(x) := \sum_{Q \in \mathcal{D}} |Q|^{\alpha/n} \langle f \rangle_Q 1_Q(x).$$

The following lemma is well–known and shows that fractional maximal and fractional integral operators can be estimated pointwise by sums of dyadic operators. For the proof of the fractional integral estimate see [20]; the proof of the estimate for the fractional maximal operator is obvious given the fact that for every cube, Q, there is a cube, P_Q in a dyadic grid such that $Q \subset P_Q$ and $|P_Q| \leqslant 3^n \, |Q|$.

Lemma 2.4. *Let* M_α *be the fractional maximal operator and* I_α *be the fractional integral operator. There is a collection of* 3^n *dyadic grids such that the following point–wise equivalences hold for all non–negative* f*:*

$$M_\alpha f \simeq \sum_{k=1}^{3^n} M_\alpha^{\mathcal{D}_k} f \quad \textit{and} \quad I_\alpha f \simeq \sum_{k=1}^{3^n} I_\alpha^{\mathcal{D}_k} f.$$

Remark 2.5. When proving the estimates below for the dyadic fractional maximal operator, it is more convenient to deal with the following truncated version:

$$1_{Q_0}(x) \sup_{Q \in \mathcal{D}: Q \subset Q_0} |Q|^{\alpha/n} \langle f \rangle_Q 1_Q(x). \tag{2.6}$$

We then prove estimates that are independent of Q_0 and appeal to the monotone convergence theorem to conclude the desired results. Assuming that f is finite almost everywhere (which will always be the case for us), we can further simplify matters. We start by building a stopping collection, $\mathcal{S}$. Initialise $\{Q_0\} \to \mathcal{S}$, and in the recursive stage, if $P \in \mathcal{S}$ is minimal, add to $\mathcal{S}$ all maximal children Q of P such that $|Q|^{\alpha/n} \langle f \rangle_Q > 4 |P|^{\alpha/n} \langle f \rangle_P$. For a cube $Q \subset Q_0$, let Q^S denote the $\mathcal{S}$–parent of Q. Similarly, let $\mathrm{ch}(S)$ denote the maximal $\mathcal{S}$–descendants of S. Finally, let $E_Q = Q \setminus \mathrm{ch}(Q)$. A simple computation shows that for every $S \in \mathcal{S}$,

$$\sum_{Q \in \mathrm{ch}(S)} |S| \leqslant \frac{1}{2} |S| \qquad \text{and} \qquad |S| \leqslant 2 |E_S| .$$

That is, the stopping collection $\mathcal{S}$ is sparse. Additionally, the E_Q are pairwise disjoint and for almost every $x \in Q_0$ there is some Q with $x \in E_Q$ (this follows from the fact that $f = \infty$ on a set of measure zero). Thus, we may further reduce (2.6) to:

$$1_{Q_0}(x) \sup_{Q \in \mathcal{D}: Q \subset Q_0} |Q|^{\alpha/n} \langle f \rangle_Q 1_Q(x) = \sum_{Q \in \mathcal{S}: Q \subset Q_0} |Q|^{\alpha/n} \langle f \rangle_Q 1_{E_Q}(x). \tag{2.7}$$

We also note that if $\{E_Q\}_{Q\in\mathcal{D}}$ is any collection of pairwise disjoint sets such that $E_Q \subset Q$, then $\sum_{Q\in\mathcal{D}} |Q|^{\alpha/n} \langle f\rangle_Q 1_{E_Q}(x) \leqslant M_\alpha f(x)$.

There is a similar reduction for the dyadic fractional integral operator. Again, we may reduce matters to:

$$1_{Q_0}(x) \sum_{Q\in\mathcal{D}:Q\subset Q_0} |Q|^{\alpha/n} \langle f\rangle_Q 1_Q(x). \tag{2.8}$$

We now create the stopping family by initialising $\{Q_0\} \to \mathcal{S}$ and in the recursive stage, if $P \in \mathcal{S}$ is minimal, add to $\mathcal{S}$ all maximal children Q of P such that $\langle f\rangle_Q > 4\langle f\rangle_P$. Note that we are stopping on *averages*, not fractional averages. Again, simple computations show that $\mathcal{S}$ is sparse. For fixed $x \in Q_0$, and fixed $S \in \mathcal{S}$, the sequence $\{|Q|^{\alpha/n} 1_Q(x)\}_{Q^s=S}$ is geometric and so

$$\sum_{Q^s=S} |Q|^{\alpha/n} 1_Q(x) \simeq C_{\alpha,n} |S|^{\alpha/n} 1_S(x). \tag{2.9}$$

Therefore, the sum in (2.8) can be estimated as:

$$\begin{aligned}\sum_{S\in\mathcal{S}} \sum_{Q^s=S} |Q|^{\alpha/n} \langle f\rangle_Q 1_Q(x) &\lesssim \sum_{S\in\mathcal{S}} \langle f\rangle_S \sum_{Q^s=S} |Q|^{\alpha/n} 1_Q(x) \\ &\lesssim \sum_{S\in\mathcal{S}} |S|^{\alpha/n} \langle f\rangle_S 1_S(x). \end{aligned}\tag{2.10}$$

Therefore, in all estimates below, for fixed f, we can replace the operator of interest with one from the right hand side of (2.7) or (2.10); our estimates will be independent of sparse collection $\mathcal{S}$ and root Q_0. $\square$

We have the following well–known testing conditions for dyadic operators, originally due to Sawyer. See [84, 36, 51].

Theorem C. *Let $1 < p \leqslant q < \infty$, let $\mathcal{D}$ be a dyadic grid and let $\mathcal{S} \subset \mathcal{D}$ be sparse. Let* T

be the operator given by $\mathsf{T}f = \sum_{Q\in\mathcal{S}} |Q|^{\alpha/n} \langle f\rangle_Q 1_Q$. *Define:*

$$\beta_1 := \sup_{P\in\mathcal{S}} \frac{1}{\sigma(P)^{q/p}} \int_P \left| \sum_{Q\in\mathcal{S}:Q\subset P} |Q|^{\alpha/n} \langle\sigma\rangle_Q 1_Q(x) \right|^q w(x)dx,$$

$$\beta_2 := \sup_{P\in\mathcal{S}} \frac{1}{w(P)^{p'/q'}} \int_P \left| \sum_{Q\in\mathcal{S}:Q\subset P} |Q|^{\alpha/n} \langle w\rangle_Q 1_Q(x) \right|^{p'} \sigma(x)dx.$$

Then:

$$\|T_\sigma : L^p(\sigma) \to L^q(w)\| \lesssim \beta_1 + \beta_2.$$

2.3 *Sparse Bilinear Forms*

2.3.1 Oscillatory Singular Integrals

Recall the notion of a *sparse* collection of cubes $\mathcal{S}$ in $\mathbb{R}^n$ that requires the existence of a set $E_Q \subset Q$ for each $Q \in \mathcal{S}$ so that (a) $|E_Q| > c|Q|$ for each $Q \in \mathcal{S}$, and (b) the collection of sets $\{E_Q : Q \in \mathcal{S}\}$ are pairwise disjoint. Here $0 < c < 1$ will be a dimensional constant that we do not track.

Definition 2.11. A *sparse* bilinear form is one of the form

$$\Lambda_{r,s}(f,g) = \sum_{Q\in\mathcal{S}} \langle f\rangle_{Q,r} \langle g\rangle_{Q,s} |Q|, \qquad 1 \leqslant r,s < \infty,$$

where $\langle f\rangle_{Q,r}^r := |3Q|^{-1} \int_{3Q} |f|^r \, dx$, and if $r = s$, then $\Lambda_r = \Lambda_{r,r}$.

We consider Calderón-Zygmund singular integral operators T, which can also be defined as L^2 bounded convolution operator (with Calderón-Zygmund kernel K) given by

$$\langle Tf, g\rangle = \iint K(x-y)f(y)g(x)\, dx\, dy$$

for compactly supported functions f, g with disjoint supports. Notable examples are $K(y) = 1/y$ in dimension one, and the Riesz transform kernels $y/|y|^{n+1}$, in dimension n.

We consider polynomials of a fixed degree d, given by

$$P(x,y) = \sum_{\alpha,\beta \,:\, |\alpha|+|\beta| \leqslant d} \lambda_{\alpha,\beta} x^\alpha y^\beta,$$

where we use the usual multi-index notation. The polynomial modulated Calderón-Zygmund operators are

$$T_P f(x) = \int e^{iP(x,y)} K(y) f(x-y) \, dy.$$

The L^p result below is a special case of the results of Ricci and Stein [78, 79], and the weak-type result is due to Chanillo and Christ [13].

Theorem D. *For $1 < p < \infty$, the operator T_P is bounded on L^p, that is*

$$\|T_P \ : \ L^p \mapsto L^p\| \lesssim 1,$$

where the implied constant depends on the degree of P, and in particular is independent of λ. Moreover, T_P maps L^1 to weak L^1, with the same bound.

The dependence on the polynomial being felt only through the degree of P is important in many applications, see [79]. This dependence continues to hold in the Theorem below, the proof of which we deffer to Section 6.1.

Theorem 2.12. *For each $1 < r < 2$, Calderón-Zygmund operator T, polynomial $P = P(y)$ of degree d and functions f, g with bounded support, there is a bilinear form Λ_r so that*

$$|\langle T_P f, g\rangle| \lesssim \Lambda_r(f,g).$$

The implied constant depends only on T, the degree d, dimension n and choice of $r > 1$.

2.3.2 Random Singular Integrals

Define a sequence of Bernoulli random variables $\{X_n \ : \ n \neq 0\}$ with $\mathbb{P}(X_n = 1) = |n|^{-\alpha}$, where $0 \leqslant \alpha < 1$. The set $\{n \ : \ X_n = 1\}$ is a.s. infinite by the Borel-Cantelli Lemma. We consider the random Hilbert transform and maximal function below:

$$H_\alpha f(x) = \sum_{n \neq 0} \frac{X_n}{n^{1-\alpha}} f(x-n),$$

$$M_\alpha f(x) = \sup_{n>0} \Big| \frac{1}{S_N} \sum_{n=1}^{N} X_n f(x-n) \Big|, \qquad \text{where} \qquad S_N = \sum_{n=1}^{N} X_n.$$

Our sparse bound here is more restrictive, with the value of the sparse index r depending on the parameter α.

Theorem 2.13. *For any $0 < \alpha < 1$, $1 + \alpha < r < 2$, the following holds almost surely: For all functions f, g finitely supported on $\mathbb{Z}$, there is a bilinear sparse operator Λ_r so that*

$$|\langle H_\alpha f, g \rangle| \lesssim \Lambda_r(f, g).$$

The same inequality holds for M_α. (The sparse operator can be taken non-random, but the implied constant is random.)

Weighted inequalities are a corollary. They are the first we know of holding for operators defined on sets of integers with zero asymptotic density. We state these corollaries in Chapter 6 and defer their proofs, along with the proof of Theorem 2.13, to Section 6.2.

CHAPTER 3

CALDERÓN-ZYGMUND OPERATORS

This chapter develops two-weight inequalities for Calderón-Zygmund operators. The work is inspired by a general question: What is the 'simplest' condition which is analogous to the Muckenhoupt A_p condition, and is sufficient for a two weight inequality to hold for all Calderón-Zygmund operators? This question arose shortly after the initial successes of the Muckenhoupt's 1972 report that the maximal function is bounded on a weighted L^p space if and only if the weight is in A_p [66]. A year later, Hunt-Muckenhoupt-Wheeden discover that the same is true of the conjugate function [34]. In both of these works, the weight in the domain and range are the same. It is natural to ask what can be done when the operators map between different weighted spaces. Nearly a decade later, Neugebauer proves a result that is fruitful in extending many one-weight inequaltites to the two-weight setting [68], which lead to the notion of testing the density of the weights in function spaces of slightly stronger norms. This theme has been investigated by many authors, with motivations coming from potential applications in different settings where Calderón-Zygmund operators appear, see for instance [88, 26] for two disparate applications. More relevant citations are in the introduction to [21], for instance.

Concerning the maximal operator itself, the finest result in this direction is due to Pérez [72]: A sharp integrability condition is used to describe a class of Orlicz spaces, and an A_p like condition, which is a sufficient condition for a two weight inequality for the maximal function. We do not recall the exact conditions, since the entropy conditions used below allow a shorter presentation of more general results. For the maximal function, this is Theorem 3.3 below.

Pérez also raised two conjectures concerning singular integrals, on being the so-called two–bump conjecture resolved in [67, 59], and the so-called separated bump conjecture which is unresolved, [21, 49].

Several recent papers have focused on the role of the A_∞ constant in completing these estimates. This theme was started in [50], and was further quantified in several papers [61, 58, 46, 40, 39, 37, 38].

Recently, Treil-Volberg [86] combined these two trends in a single approach, which they termed the *entropy bounds*, and as is explained n [86, § 2], this approach yields (slightly) stronger results than that of the Orlicz function approach. In what follows, we will extend their results to the L^p-setting, using very short proofs.

3.1 Definition

We say that $K : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ is a *Calderón-Zygmund kernel* if for some constants and $C_K > 0$, and $0 < \eta < 1$, such that these conditions hold: For $x, x', y \in \mathbb{R}^d$

$$
\begin{aligned}
\|K(\cdot,\cdot)\|_\infty &< \infty, \\
|K(x,y)| &< C_K |x-y|^{-d}, \qquad x \neq y, \\
|K(x,y) - K(x',y)| &< C_K \frac{|x-x'|^\eta}{|x-y|^{d+\eta}}, \qquad \text{if } 2|x-x'| < |x-y|,
\end{aligned}
$$

and a fourth condition, with the roles of the first and second coordinates of $K(x,y)$ reversed also holds. These are typical conditions, although in the first condition, we have effectively truncated the kernel, at the diagonal and infinity. The effect of this is that we needn't be concerned with principal values.

Given a Calderón-Zygmund kernel K as above, we can define

$$
Tf(x) := \int K(x,y) f(y)\, dy
$$

which is defined for all $f \in L^2$ and $x \in \mathbb{R}^d$. We say that T is a *Calderón-Zygmund operator*, since it necessarily extends to a bounded operator on $L^2(\mathbb{R}^d)$. We define

$$C_T := C_K + \|T : L^2 \to L^2\|. \tag{3.1}$$

It is well-known that T is also bounded on L^p, $1 < p < \infty$, with norm controlled by C_T.

3.2 Main Results

Throughout, let

$$\rho_\sigma(Q) = \frac{\int_Q M(\sigma \mathbf{1}_Q)\, dx}{\sigma(Q)}, \qquad \rho_{\sigma,\varepsilon}(Q) = \rho_\sigma(Q)\varepsilon(\rho_\sigma(Q)),$$

where ε will be an increasing function on $[1, \infty)$. But, if the role of the weight σ is understood, it is suppressed in the notation. Define

$$\lceil \sigma, w \rceil_{p,\varepsilon} := \sup_{Q \text{ a cube}} \rho_{\sigma,\varepsilon}(Q) \langle \sigma \rangle_Q^{p-1} \langle w \rangle_Q. \tag{3.2}$$

Throughout, $\langle f \rangle_Q = |Q|^{-1} \int_Q f(x)\, dx$. In this Theorem, we extend the result of Pérez [72] for the Hardy-Little-wood maximal function, denoted M, to the entropy language.

Theorem 3.3. *Let σ and w be two weights with densities, and $1 < p < \infty$. Let ε be a monotonic increasing function on $(1, \infty)$ which satisfies $\int_1^\infty \frac{dt}{\varepsilon(t)t} = 1$. Denote by $M_\sigma f = M(\sigma f)$. The following two-weight inequality holds:*

$$\|M_\sigma : L^p(\sigma) \mapsto L^p(w)\| \lesssim \lceil \sigma, w \rceil_{p,\varepsilon}^{1/p}. \tag{3.4}$$

As above, we use the notation $M_\sigma f = M(\sigma f)$ so that inequalities are stated in a self-dual way. It is natural to include 3.3 in this chapter since it is well-known that the maximal function serves as a bounding operator for the Calderón-Zygmund operators in an intuitive sense. For the most compelling result, see the famous Coifman-Fefferman inequality in [15], which says the maximal function pointwise bounds the maximal Calderón-Zygmund operators on any A_∞-weighted space. However, the dependencies of the implied constants in that inequality are delicate, so as is the case here, it is often necessary to examine the maximal and Calderón-Zygmund operators independently.

Concerning Calderón-Zygmund operators, the case of $p = 2$ below is [86, Thm 2.5]. It is slightly stronger than the two–bump results in [67, 59].

Theorem 3.5. *Let* σ *and* w *be two weights with densities, and* $1 < p < \infty$*. Let* ε *be a monotonic increasing function on* $(1, \infty)$ *which satisfies* $\int_1^\infty \frac{dt}{\varepsilon(t)t} = 1$*. Define*

$$\lfloor \sigma, w \rfloor_p := \sup_{Q \text{ a cube}} \langle \sigma \rangle_Q^{p-1} \rho_{\sigma,\varepsilon}(Q) \langle w \rangle_Q \rho_{w,\varepsilon}(Q)^{p-1} \tag{3.6}$$

For any Calderón-Zygmund operator, there holds

$$\|T_\sigma \,:\, L^p(\sigma) \to L^p(w)\| \lesssim C_T \lfloor \sigma, w \rfloor_p^{1/p} \|f\|_{L^p(\sigma)}.$$

The constant C_T *is defined in* (3.1)*.*

In the condition (3.6) above, both of the weights σ and w are 'bumped.' Below, the bump is applied to each weight separately, hence the name *separated bump* condition. The case $p = 2$ below corresponds to [86, Thm 2.6] It is slightly stronger than the corresponding results proved in [49].

Theorem 3.7. *Let* σ *and* w *be two weights with densities, and* $1 < p < \infty$*. Let* $\varepsilon_p, \varepsilon_{p'}$ *be two monotonic increasing functions on* $(1, \infty)$ *which satisfy* $\int_1^\infty \varepsilon_p(t)^{-1/p} \frac{dt}{t} = 1$*, and* *similarly for $\varepsilon_{p'}$ with root $1/p'$. For any Calderón-Zygmund operator, there holds*

$$\|T_\sigma \,:\, L^p(\sigma) \to L^p(w)\| \lesssim C_T\{\lceil \sigma, w \rceil_{p,\varepsilon_p}^{1/p} + \lceil w, \sigma \rceil_{p',\varepsilon_{p'}}^{1/p'}\}.$$

The terms involving the weights is defined in (3.2)*, and the constant C_T is defined in* (3.1).

One should not fail to note that the integrability condition imposed on $\varepsilon_p(t)^{-1}$ is stronger than in Theorem 3.5. It is not known if the condition in Theorem 3.7 is the sharp. The following result is another separated bump condition for Calderón-Zygmund Operators.

The type of theorems we are proving are known as "bumps" because they slightly strengthen the joint A_p characteristic. The bumps in Theorem 3.7 were introduced in [86] and are known as "entropy bumps". However, the bumps in Theorem 3.8 are slightly different due to their dependence on the behavior of α_p at zero, and they seem to be new. There is a long history of theorems of this type (see for example [20, 22, 17, 21, 49, 37, 61, 68, 67, 72]), but in [86] it is shown that under some mild conditions, the entropy bumps are smaller than other bumps, encouraging progress with this approach.

Our proof builds on the techniques in the proofs of Theorem 3.7 and uses an interesting formula by Hytönen in [36] that generalizes the expansion of sums like $(\sum_j a_j)^2$ to powers other than 2. This formula is powerful and it seems to have been first observed in [36].

Theorem 3.8. *Let σ and w be two weights with densities, and $1 < p < \infty$. Define*

$$[[\sigma, w]]_{p,\alpha_p} := \sup_{Q \text{ a cube}} \langle w \rangle_Q \langle \sigma \rangle_Q^{p-1} \alpha_p(\langle \sigma \rangle_Q),$$

where α_p is a function that is decreasing on $(0,1)$ and increasing on $(1,\infty)$ and that *satisfies* $\sum_{r\in\mathbb{Z}} \alpha_p(2^{-r})^{-\frac{1}{p}} < \infty$. *Then it follows that*

$$\|T_{\mathcal{S}}\sigma\cdot : L^p(\sigma) \to L^p(w)\| \lesssim [[\sigma, w]]_{p,\alpha_p}^{\frac{1}{p}} + [[w, \sigma]]_{p',\alpha_{p'}}^{\frac{1}{p'}}.$$

For all of the above, the method of proof we use is, like Lerner [57], to reduce to sparse operators. With the recent argument of Lacey [47], this reduction now applies more broadly, namely it applies to (a) Calderón-Zygmund operators on Euclidean spaces as stated above; (b) non-homogenous Calderón-Zygmund operators; and (c) general martingales.

After the reduction to sparse operators, we use arguments involving pigeonholes, stopping times, reduction to testing conditions, and an A_p-A_∞ inequality. These are the shortest proofs we could find.

3.3 An Entropy Condition for the Maximal Function

We prove the maximal function estimate (3.4). It suffices to prove the theorem with the maximal function replaced by a dyadic version, since it is a classical fact that in dimension d, there are at most 3^d choices of shifted dyadic grids $\mathcal{D}_j$, for $1 \leqslant j \leqslant 3^d$, which approximate any cube in $\mathbb{R}^d$.

By Sawyer's characterization [83] of the two weight maximal function inequality, it suffices to check that inequality for $f = \mathbf{1}_{Q_0}$, and any dyadic cube Q_0. Namely, we should prove

$$\int_{Q_0} M(\sigma \mathbf{1}_{Q_0})^p \, dw \lesssim \lceil \sigma, w \rceil_{p,\varepsilon} \sigma(Q_0).$$

To do so, let $\mathcal{S}$ be a sequence of stopping cubes for σ, defined as follows. The root of $\mathcal{S}$ is Q_0, and if $S \in \mathcal{S}$, the maximal dyadic cubes $Q \subset S$ such that $\langle\sigma\rangle_Q > 4\langle\sigma\rangle_S$ are also in $\mathcal{S}$. Note that this is a sparse collection of cubes. Then, we have

$$\mathbf{1}_{Q_0} \cdot M(\sigma \mathbf{1}_{Q_0}) \lesssim \sum_{S\in\mathcal{S}} \langle\sigma\rangle_S \mathbf{1}_{E_S}$$

where $E_S := S \setminus \bigcup\{S' \in \mathcal{S} \ : \ S' \subsetneq S\}$. The collection $\mathcal{S}$ is sparse, and the sets E_S are pairwise disjoint, hence,

$$\int_{Q_0} M(\sigma \mathbf{1}_{Q_0})^p \, dw \lesssim \sum_{S \in \mathcal{S}} \langle \sigma \rangle_S^p w(S).$$

The sparse collection $\mathcal{S}$ is divided into collations $\mathcal{S}_{a,r}$, for $a \in \mathbb{Z}$ and $r \in \mathbb{N}$ defined by $S \in \mathcal{S}_{a,r}$ if and only if

$$2^a \sim \langle \sigma \rangle_S^{p-1} \langle w \rangle_Q \rho_{\sigma,\varepsilon}(Q), \quad \text{and} \quad 2^r \sim \rho(Q).$$

Notice that $\mathcal{S}_{a,r}$ is empty if $\lceil \sigma, w \rceil_{p,\varepsilon} < 2^{a-1}$.

Holding a and r constant, it follows that

$$\sum_{S \in \mathcal{S}_{a,r}} \langle \sigma \rangle_S^p w(S) \lesssim 2^a \sum_{S \in \mathcal{S}_{a,r}} \frac{\sigma(S)}{2^r \varepsilon(2^r)} \lesssim 2^a \sum_{\text{maximal } S \, \in \, \mathcal{S}_{a,r}} \frac{\int_S M(\sigma \mathbf{1}_S)}{2^r \varepsilon(2^r)} \lesssim 2^a \frac{\sigma(Q_0)}{\varepsilon(2^r)}.$$

Notice that sparsity is essential to the domination of the sum by the maximal function in the second line. To sum this over $r \in \mathbb{N}$, we need the integrability condition $\int_1^\infty \frac{dt}{\varepsilon(t)t} = 1$. Take pth roots and sum over appropriate $a \in \mathbb{Z}$ to conclude.

3.4 A Two–Bump Condition

This section is dedicated to the proof of the two–bump inequality that is Theorem 3.5. Fix a sparse collection $\mathcal{S}$ so that for all cubes $Q \in \mathcal{S}$ there holds, for some $a \in \mathbb{Z}$,

$$2^a \sim \langle \sigma \rangle_Q^{p-1} \rho_{\sigma,\varepsilon}(Q) \langle w \rangle_Q \rho_{\sigma,\varepsilon_{p'}}(Q)^{p-1}$$

Here, $2^{a-1} \leqslant \lfloor \sigma, w \rfloor_p$. In this case, we will verify that the norm of the associated sparse operator is bounded as by $\lesssim 2^{a/p}$. This estimate is clearly suitable in relevant $a \in \mathbb{Z}$.

The proof is by duality. Thus, for $f \in L^p(\sigma)$ and $g \in L^{p'}(w)$, we bound the pairing $\langle S(\sigma f), gw\rangle$. In so doing, we will write

$$\langle f\sigma\rangle_Q = \langle f\rangle_Q^\sigma \langle \sigma\rangle_Q,$$

where $\langle f\rangle_Q^\sigma$ is the average of f relative to weight σ on the cube Q. Then,

$$\begin{aligned} 2^{-a/p}\langle S(\sigma f), gw\rangle &= 2^{-a/p}\sum_{Q\in\mathcal{S}} \langle \sigma f\rangle_Q \langle gw\rangle_Q \cdot |Q| \\ &= \sum_{Q\in\mathcal{S}} \langle f\rangle_Q^\sigma \langle\sigma\rangle_Q^{1/p}\left\{\frac{\langle\sigma\rangle_Q^{1/p'}\langle w\rangle_Q^{1/p}}{2^{a/p}}\right\}\langle w\rangle_Q^{1/p'}\langle g\rangle_Q^w \cdot |Q| \\ &\lesssim \sum_{Q\in\mathcal{S}} \langle f\rangle_Q^\sigma \frac{\sigma(Q)^{1/p}}{\rho_{\sigma,\varepsilon}(Q)^{1/p}} \cdot \langle g\rangle_Q^w \frac{w(Q)^{1/p'}}{\rho_{w,\varepsilon}(Q)^{1/p'}}. \end{aligned}$$

Apply Hölder's inequality to the last expression. It clearly suffices to show that

$$\sum_{Q\in\mathcal{S}} (\langle f\rangle_Q^\sigma)^p \frac{\sigma(Q)}{\rho_\sigma(Q)} \lesssim \|f\|_{L^p(\sigma)}^p,$$

and similarly for g.

This last expression is a Carleson embedding inequality. It is well known that it suffices to check this inequality for $f = \mathbf{1}_{Q_0}$, for $Q_0 \in \mathcal{S}$, and the assumption that Q_0 is the maximal element in $\mathcal{S}$. But notice that the sum to control is then

$$\begin{aligned} \sum_{Q\in\mathcal{S}} \frac{\sigma(Q)}{\rho_\sigma(Q)} &\lesssim \sum_{r=1}^\infty \sum_{\substack{Q\in\mathcal{S} \\ \rho_\sigma(Q)\sim 2^r}} \frac{\sigma(Q)}{2^r\varepsilon(2^r)} \\ &\lesssim \sum_{r=1}^\infty \sum_{\substack{Q \text{ maximal s.t.} \\ Q\in\mathcal{S},\ \rho_\sigma(Q)\sim 2^r}} \frac{\int_Q M(\sigma\mathbf{1}_Q)\, dx}{2^r\varepsilon(2^r)} \\ &\lesssim \sigma(Q_0)\sum_{r=0}^\infty \frac{1}{\varepsilon(2^r)}. \end{aligned}$$

The middle inequality follows from sparseness. The last sum over r should be finite, which is integrability condition $\int_1^\infty \frac{dt}{t\varepsilon(t)} = 1$. The proof is complete.

3.5 Separated Bump Condition I

This section is dedicated to the proof of Theorem 3.7. In fact, we will follow the argument presented here with an alternative one. There are two key preliminaries in the first proof. One is the testing condition Theorem A presented in Chapter 2. Namely, it suffices to verify: For any dyadic cube Q_0,

$$\int_{Q_0} \Big| \sum_{Q \in \mathcal{S} \,:\, Q \subset Q_0} \langle \sigma \rangle_Q \mathbf{1}_Q \Big|^p \, dw \lesssim \lceil \sigma, w \rceil_{p,\varepsilon_p} \sigma(Q_0).$$

The dual inequality will also hold, and so complete the proof of Theorem 3.7.

The other ingredient is following Lemma 3.9 below. In the current setting, it originates in [46], though we give a more convenient reference below. Notice that the bound on the right in the estimates below are specific to the sparse collection being used.

Lemma 3.9. *[37, Prop. 5.3] Let $\mathcal{S}$ be a sparse collection of cubes all contained in a cube Q_0, defining a sparse operator S. For two weights σ and w,*

$$\int_{Q_0} (S\sigma \mathbf{1}_{Q_0})^p \, dw \lesssim A_p(\mathcal{S}) A_\infty(\mathcal{S}) \sigma(Q_0), \tag{3.10}$$

where $A_p(\mathcal{S}) := \sup_{Q \in \mathcal{S}} \langle \sigma \rangle_Q^{p-1} \langle w \rangle_Q$ *and* $A_\infty(\mathcal{S}) := \sup_{Q \in \mathcal{S}} \frac{\int_Q M(\mathbf{1}_Q \sigma) \, dx}{\sigma(Q)}$.

For integers $a \in \mathbb{Z}$, and $r \in \mathbb{N}$ set $\mathcal{S}_{a,r}$ to be all those cubes $Q \in \mathcal{S}$ such that $Q \subset Q_0$,

$$2^a \sim \rho_{\sigma,\varepsilon_p}(Q) \langle \sigma \rangle_Q^{p-1} \langle w \rangle_Q, \qquad \text{and} \qquad 2^r \sim \frac{\int_Q M(\mathbf{1}_Q \sigma) \, dx}{\sigma(Q)}.$$

This collection is empty if $\lceil \sigma, w \rceil_{p,\varepsilon_p} < 2^{a+1}$. By construction, $A_\infty(\mathcal{S}_{a,r}) \lesssim 2^r$, and

$$A_p(\mathcal{S}_{a,r}) \lesssim \frac{2^a}{\rho_{\sigma,\varepsilon_p}(Q)} \simeq \frac{2^a}{2^r \varepsilon_p(2^r)}.$$

Thus, from (3.10), we have

$$\int_{Q_0} \left[\sum_{Q \in \mathcal{S}_{a,r}} \langle \sigma \rangle_Q \mathbf{1}_Q \right]^p dw \lesssim A_p(\mathcal{S}_{a,r}) A_\infty(\mathcal{S}_{a,r}) \sigma(Q_0) \lesssim \frac{2^a}{\varepsilon_p(2^r)} \sigma(Q_0).$$

Take pth root, and sum over the relevant $a \in \mathbb{Z}$, and $r \in \mathbb{N}$. The sum over r is finite since $\int_1^\infty \frac{dt}{t\varepsilon_p(t)^{1/p}} = 1$, completing the proof.

3.5.1 An Alternative Proof

We first give the proof of Theorem 3.7 in the case $p = 2$. We will verify the testing conditions hold; we will only verify the first condition as the second condition is verified similarly. Fix $P \in \mathcal{S}$. By the triangle inequality and the summability condition of ε_2, it suffices to show

$$\int_P \left| \sum_{Q \in \mathcal{Q}_r} \langle \sigma \rangle_Q 1_Q \right|^2 w \lesssim \frac{1}{\varepsilon_2(2^r)} [\sigma, w]_{2,\varepsilon_2} \sigma(P), \tag{3.11}$$

where $\mathcal{Q}_r := \{Q : Q \subset P \text{ and } \rho_\sigma(Q) \simeq 2^r\}$ for $r \in \mathbb{N}$. Since two cubes in $\mathcal{Q}_r$ are either nested or disjoint, there holds

$$\left| \sum_{Q \in \mathcal{Q}_r} \langle \sigma \rangle_Q 1_Q(x) \right|^2 \simeq \sum_{Q \in \mathcal{Q}_r} \sum_{Q' \subset Q} \langle \sigma \rangle_Q \langle \sigma \rangle_{Q'} 1_{Q'}(x).$$

Inserting this into (3.11), and using $\rho_\sigma(Q) \simeq 2^r$ for $Q \in \mathcal{Q}_r$,

$$\begin{aligned}\int_P \left| \sum_{Q \in \mathcal{Q}_r} \langle \sigma \rangle_Q 1_Q \right|^2 w &\simeq \sum_{Q \in \mathcal{Q}_r} \sum_{Q' \subset Q} \langle \sigma \rangle_Q \langle \sigma \rangle_{Q'} w(Q') \\ &= \sum_{Q \in \mathcal{Q}_r} \langle \sigma \rangle_Q \sum_{Q' \subset Q} |Q'| \, \langle \sigma \rangle_{Q'} \langle w \rangle_{Q'} \frac{\rho_\sigma(Q)\varepsilon(\rho_\sigma(Q))}{\rho_\sigma(Q)\varepsilon(\rho_\sigma(Q))}. \\ &\lesssim \frac{1}{2^r \varepsilon_2(2^r)} [\sigma, w]_{2,\varepsilon_2} \sum_{Q \in \mathcal{Q}_r} \langle \sigma \rangle_Q \sum_{Q' \subset Q} |Q'| \,. \end{aligned}$$

Since $\mathcal{Q}_r$ is sparse, $\sum_{Q \in \mathcal{Q}_r} \langle \sigma \rangle_Q \sum_{Q' \subset Q} |Q'| \lesssim \sum_{Q \in \mathcal{Q}_r} \sigma(Q)$.

Recall that for a sparse collection $\mathcal{S}$ of cubes, the following holds uniformly over $P \in \mathcal{S}$: $|\cup_{Q \in \mathcal{S} : Q \subset P} Q| \leqslant \frac{1}{2} |P|$. This implies that the following holds uniformly over all $P \in \mathcal{S}$: $\sum_{Q \in \mathcal{S} : Q \subset P} |Q| \lesssim |P|$. For a cube $Q \in \mathcal{S}$, let $E_Q := Q \setminus \cup_{S \in \mathcal{S} : S \subset Q} S$ and note that $|E_Q| \simeq |Q|$. Set $\mathcal{Q}_r^*$ to be the maximal cubes in $\mathcal{Q}_r$. Using the fact that $|E_Q| \simeq |Q|$ and that $\{E_Q\}$ are pairwise disjoint, it follows that:

$$\begin{aligned}\sum_{Q \in \mathcal{Q}_r} \sigma(Q) &\simeq \sum_{Q^* \in \mathcal{Q}_r^*} \int_{Q^*} \sum_{Q \subset Q^*} \langle \sigma \rangle_Q 1_{E_Q} \\ &\leqslant \sum_{Q^* \in \mathcal{Q}_r^*} \int_{Q^*} M(\sigma 1_{Q^*}) \\ &\leqslant 2^r \sum_{Q^* \in \mathcal{Q}_r^*} \sigma(Q^*). \end{aligned} \tag{3.12}$$

Since the cubes in $\mathcal{Q}_r^*$ are pairwise disjoint, the sum is bounded by $\sigma(P)$, as desired.

To use a similar idea for $p \neq 2$ we need the following theorem proven in [36].

Theorem E. *Let $\mathcal{Q}$ be any collection of cubes. With obvious notation, there holds*

$$\int_P \left(\sum_{Q \in \mathcal{Q} : Q \subset P} \langle \sigma \rangle_Q 1_Q \right)^p w \lesssim [w, \sigma]_p^{\mathcal{Q}} \sum_{Q \subset P} \langle \sigma \rangle_Q |Q| \,.$$

We use this to prove Theorem 3.7 for all $p > 1$. We will verify the first testing condition in A, and the dual condition is verified similarly. Thus, let P be any cube in $\mathcal{S}$. For $r \geqslant 0$, let $\mathcal{Q}_r = \{Q \subset P : \rho_\sigma(Q) \simeq 2^r\}$. Note that for these cubes, there holds

$$[w,\sigma]_p^{\mathcal{Q}} \lesssim \frac{1}{2^r \varepsilon_p(2^r)} [w,\sigma]_{p,\varepsilon_p}.$$

Therefore, by the triangle inequality and Lemma E,

$$\begin{aligned}
\left(\int_P \left(\sum_{Q:Q\subset P} \langle\sigma\rangle_Q 1_Q\right)^p w\right)^{\frac{1}{p}} &\leqslant \sum_{r\geqslant 0} \left(\int_P \left(\sum_{Q\in\mathcal{Q}_r} \langle\sigma\rangle_Q 1_Q\right)^p w\right)^{\frac{1}{p}} \\
&\lesssim [w,\sigma]_{p,\varepsilon_p}^{\frac{1}{p}} \sum_{r\geqslant 0} \frac{1}{\varepsilon_p(2^r)^{\frac{1}{p}}} \left(\frac{1}{2^r} \sum_{Q\in\mathcal{Q}_r} \sigma(Q)\right)^{\frac{1}{p}} \\
&\lesssim [w,\sigma]_{p,\varepsilon_p}^{\frac{1}{p}} \sum_{r\geqslant 0} \frac{1}{\varepsilon_p(2^r)^{\frac{1}{p}}} \sigma(P)^{\frac{1}{p}}.
\end{aligned}$$

In the last estimate, we used the fact that for the cubes in $\mathcal{Q}_r$, $\rho_\sigma(Q) \simeq 2^r$ and so we can use the same estimate as in (3.12). The summability condition on ε_p completes the proof.

3.6 Separated Bump Condition II

We conclude the chapter with a proof of the second separated bump condition mentioned: Theorem 3.8. As above, it suffices to verify the testing conditions in Theorem A, and we will only verify the first. Thus, let P be any cube in $\mathcal{S}$. For $r \in \mathbb{Z}$ let $\mathcal{Q}_r = \{Q \subset P : \langle\sigma\rangle_Q \simeq 2^r\}$. Using the summability condition on α_p, as in the proof of Theorem 3.7, we may assume that all cubes are contained in $\mathcal{Q}_r$.

Again let $\mathcal{Q}_r^*$ denote the maximal cubes in $\mathcal{Q}_r$. Using Lemma **??**, there holds

$$\begin{aligned}
\int_P \left(\sum_{Q \in \mathcal{Q}_r} \langle \sigma \rangle_Q 1_Q \right)^p w &\lesssim \frac{1}{\alpha_p(2^r)} [[w, \sigma]]_{p,\alpha_p} \sum_{Q \in \mathcal{Q}_r} \langle \sigma \rangle_Q |Q| \\
&\simeq \frac{1}{\alpha_p(2^r)} [[w, \sigma]]_{p,\alpha_p} \sum_{Q^* \in \mathcal{Q}_r^*} \sum_{Q \subset Q^*} |Q| \\
&\simeq \frac{1}{\alpha_p(2^r)} [[w, \sigma]]_{p,\alpha_p} \sum_{Q^* \in \mathcal{Q}_r^*} |Q|^* .
\end{aligned}$$

In the second line we used the definition of $\mathcal{Q}_r$ and in the third line we used sparseness. Again, using the definition of $\mathcal{Q}_r$, the sum is equivalent to $\sum_{Q^* \in \mathcal{Q}_r^*} \sigma(Q^*)$ and by the maximality of the cubes in $\mathcal{Q}^*$, it follows that this sum is dominated by $\sigma(P)$.

CHAPTER 4

FRACTIONAL INTEGRAL OPERATORS

We are concerned with two-weight inequalities for the fractional maximal and fractional integral operators. The goal is to find simple, A_p-like conditions for a pair of weights (non–negative, locally integrable functions) σ, w to ensure

$$\|T^{\sigma} : L^p(\sigma) \to L^q(w)\| < \infty, \tag{4.1}$$

where T denotes a fractional maximal or fractional integral operator, and $T^{\sigma}(f) := T(\sigma f)$. One popular approach, initiated by Neugebauer in [68] and developed by Pérez in [71, 70], has been to slightly strengthen the A_p characteristic by introducing new factors. These new factors, known as bumps, have come in different forms. For example, Neugebauer requires that the weights $\sigma^{1+\epsilon}$ and $w^{1+\epsilon}$ belong to A_p, while Pérez requires that the two weights have finite Orlicz norm. The Orlicz approach is also taken by Cruz-Uribe and Moen in [20]. See the recent paper of Cruz–Uribe [18] and the references therein for more information.

In the context of Calderón–Zygmund operators, Treil–Volberg have recently introduced the notion of *entropy bounds* and are able to deduce stronger results than have been obtained using the Orlicz approach [86]. Lacey and the author [52] simplified and extended the approach to the entropy conditions in the singular integral case. We use these same techniques to prove similar results for the fractional integral and fractional maximal operators. These results represent an extension of what is known, and can be proved by relatively simple techniques. In particular, we require that our weights satisfy certain bump or separated bump conditions (to be defined below.) It is not known to what extent these results are sharp. However, Treil and Volberg show that the bumps used here are - in many cases of interest - smaller than the Orlicz–based bumps.

Before stating the main theorems, we give some definitions. For $0 < \alpha < n$, the fractional maximal operator for functions defined on $\mathbb{R}^n$ is

$$M_\alpha f(x) := \sup_{Q \text{ a cube}} \frac{1_Q(x)}{|Q|^{1-\frac{\alpha}{n}}} \int_Q |f(y)|\, dy,$$

and the fractional integral operator is

$$I_\alpha f(x) := \int_{\mathbb{R}^n} \frac{f(y)}{|y-x|^{n-\alpha}} dy.$$

4.1 Main Results

One reasonable generalization of the Muckenhoupt A_p condition to the present setting is to set $[\sigma, w] := \sup_{\text{cubes } Q} \sigma(Q)^{1/p'} w(Q)^{1/q} |Q|^{\alpha/(n-1)}$. Ideally, we would like for (4.1) to hold when $[\sigma, w]$ is finite. This condition is insufficient (see [19] for a counter example in the case of the fractional maximal operator). This condition *is* enough, however, to deduce *weak-type* bounds for the maximal operator. We present an alternate proof of this well–known result in Section 4.2 as an example of the techniques used in the main theorems of this section; see [19] for another proof. In particular, there holds:

Theorem 4.2. *With* $[\sigma, w]$ *defined as above,* M_α *the fractional maximal operator, and* $1 \leqslant p \leqslant q \leqslant \infty$, *there holds:*

$$\|M_\alpha(\sigma \cdot) : L^p(\sigma) \to L^{q,\infty}(w)\| \lesssim [\sigma, w].$$

Since the finiteness of $[\sigma, w]$ is not enough to deduce strong bounds, we use two types of bumped conditions to deduce the strong estimates. The first set of conditions on the weights that we consider require a single bump (compare with the separated bumps to be discussed later). Set $\rho_\sigma(Q) := \frac{1}{\sigma(Q)}\int_Q M(\sigma 1_Q)$, and define ρ_w similarly, where M is the Hardy–Littlewood maximal operator. We deal first with the fractional maximal operator. In [71, 70], Pérez establishes bump conditions related to Theorems 4.3 and 4.4 using Orlicz norms.

Theorem 4.3. *Let σ and w be two weights, $1 < p \leqslant q < \infty$, and M_α be the fractional maximal operator. Let ϵ_q be a monotonic increasing function on $(1, \infty)$ that satisfies $\int_1^\infty \frac{dt}{t\epsilon_q^q(t)} = 1$. Define*

$$\beta(Q) := \frac{\sigma(Q)^{1/p'} w(Q)^{1/q}}{|Q|^{1-\alpha/n}} \rho_\sigma^{1/p}(Q)\epsilon_q(\rho_\sigma(Q)),$$

and set $\lceil \sigma, w \rceil := \sup_{Q\in\mathcal{D}} \beta(Q)$. Then

$$\|M_\alpha(f\sigma)\|_{L^q(w)} \lesssim \lceil \sigma, w \rceil \|f\|_{L^p(\sigma)}.$$

The corresponding theorem for the fractional integral operator is:

Theorem 4.4. *Let $1 \leqslant p \leqslant \infty$ and σ and w be two weights and let I_α be the fractional integral operator. Let ϵ_p be a monotonic increasing function on $(1, \infty)$ that satisfies $\int_1^\infty \frac{dt}{t\epsilon_p^p(t)} = 1$ and similarly for $\epsilon_{q'}$. Define:*

$$\beta(Q) := \frac{\sigma(Q)^{1/p'} w(Q)^{1/q}}{|Q|^{1-\alpha/n}} \rho_\sigma(Q)^{1/p}\epsilon_p(\rho_\sigma(Q))\rho_w(Q)^{1/q'}\epsilon_{q'}(\rho_w(Q)),$$

and set $\lfloor \sigma, w \rfloor := \sup_{Q\in\mathcal{Q}} \beta(Q)$. Then

$$\|I_\alpha(f\sigma)\|_{L^q(w)} \lesssim C_{\alpha,n}\lfloor \sigma, w \rfloor \|f\|_{L^p(\sigma)},$$

where $C_{\alpha,n}$ is from (2.9).

The condition in the next theorem is called a "separated bump" for obvious reasons. We use a bump defined in terms of the fractional maximal operator, namely

$$\rho_\sigma^{\alpha,p,q}(Q) := \frac{\int_Q M_\alpha(1_Q\sigma)^{q/p}\,dx}{\sigma(Q)^{q/p}},$$

or simply ρ_σ or ρ when clear. We have the following

Theorem 4.5. *Let σ and w be weights with densities, $1 < p \leqslant q < \infty$, and $\varepsilon_q, \varepsilon_{p'} : \mathbb{R}^+ \to \mathbb{R}$ be nonincreasing on $(0,1)$ and nondecreasing on $(1,\infty)$ such that $\int_0^\infty \frac{dt}{t\varepsilon_q^{1/q}(t)}$ and $\int_0^\infty \frac{dt}{t\varepsilon_{p'}^{1/p'}(t)}$ are finite. Define*

$$[[\sigma,w]]_{\alpha,p,q} := \sup_{Q \text{ a cube}} \left(|Q|^{\alpha/n}\langle\sigma\rangle_Q\right)^{q/p'} \langle w\rangle_Q \rho_\sigma^{\alpha,p,q}(Q)\varepsilon_q\left(\rho_\sigma^{\alpha,p,q}(Q)\right).$$

There holds:

$$\|I_\alpha^\sigma : L^p(\sigma) \to L^q(w)\| \lesssim C_{\alpha,n}\left([[\sigma,w]]_{\alpha,p,q}^{1/q} + [[w,\sigma]]_{\alpha,q',p'}^{1/p'}\right).$$

In Section 4.2, we give some preliminary information and lemmas that will be used below. In Section 4.3, we give a proof of the weak estimates. Section 4.4 and Section 4.5 contain the proofs of the one–bump theorems for the fractional maximal and fractional integral operators. The proofs in these sections use the theory of sparse operators discussed in Chapter 2, but avoid the explicit use of testing inequalities. Finally, Section 4.6 contains the proof of the separated bump theorem for the fractional integral operator. The proof uses both sparse operators and testing inequalities but is still elementary.

4.2 Preliminaries

In this section, we state a definition and believed-to-be-well-known theorem that will be usefull later in the chapter. We include the proof for completeness.

Definition 4.6. Given a measure μ on $\mathbb{R}^n$ and a dyadic grid, $\mathcal{D}$, a sequence of positive numbers, $\{a_Q\}_{Q\in\mathcal{D}}$, is called a p, q–*Carleson Sequence* if for every $P \in \mathcal{D}$,

$$\frac{1}{\mu(P)^{q/p}} \sum_{Q\in\mathcal{D}:Q\subset P} a_Q \lesssim 1. \tag{4.7}$$

The following is a variant of a Carleson Embedding Theorem. We are certain that Theorem F is contained in a paper, but we have not been able to find a reference. For the "continuous" version of this theorem, see [29].

Theorem F. *Let μ be a measure on $\mathbb{R}^n$, $\mathcal{D}$ be a dyadic grid, and $\{a_Q\}_{Q\in\mathcal{D}}$ be a p, q–Carleson Sequence. If $1 < p \leqslant q < \infty$, there holds:*

$$\sum_{Q\in\mathcal{D}} a_Q \left(\langle f\rangle_Q^\mu\right)^q \lesssim \|f\|_{L^p(\mu)}^q,$$

where the implied constant depends on p, q and the best constant in (4.7).

Proof. We will treat $\mathcal{D}$ as a discrete measure space with measure ν where $\nu(Q) = a_Q$. We show that the operator T with rule $(Tf)(Q) = \langle f\rangle_Q^\mu$ satisfies $\|Tf\|_{L^q(\nu)}^q \lesssim \|f\|_{L^p(\mu)}^q$. The objective then is to show that for every $\lambda > 0$, there holds:

$$\lambda^q \nu(\{Tf > \lambda\}) \lesssim (\lambda^p \mu(Mf > \lambda))^{q/p}, \tag{4.8}$$

where M is the dyadic maximal function. The lemma follows from (4.8) since the dyadic maximal function is bounded for $p > 1$:

$$\|Tf\|_{L^q(\nu)}^q \simeq \sum_{k\in\mathbb{Z}} 2^{kq}\nu(\{Tf > 2^k\}) \lesssim \left(\sum_{k\in\mathbb{Z}} 2^{kp}\mu(\{Mf > 2^k\})\right)^{q/p} \simeq \|Mf\|_{L^p(\mu)}^{q/p}.$$

We now turn to proving (4.8). Fix $\lambda > 0$, and let $\mathcal{D}_\lambda$ be the maximal elements $Q \in \mathcal{D}$ such that $\langle f\rangle_Q^\mu > \lambda$ (such maximal cubes exist since $f \in L^p(\mu)$). Using the Carleson property of the sequence $\{a_Q\}_{Q\in\mathcal{D}}$, there holds:

$$\lambda^q\nu(\{Tf > \lambda\}) = \lambda^q \sum_{P\in\mathcal{D}_\lambda} \sum_{Q\in\mathcal{D}_\lambda : Q\subset P} a_Q \leqslant \sum_{P\in\mathcal{D}_\lambda} (\lambda^p\mu(P))^{q/p} \leqslant (\lambda^p\mu(\{Mf > \lambda\}))^{q/p}.$$

The last inequality follows by the disjointness of the $P \in \mathcal{D}_\lambda$ and the fact that $q/p \geqslant 1$. $\square$

4.3 A Weak-Type Inequality for the Fractional Maximal Operator

By Lemma 2.4, Theorem 4.2 follows from the following lemma.

Lemma 4.9. *Let $1 \leqslant p \leqslant q < \infty$ and σ and w be two weights. Let $\mathcal{D}$ be a dyadic grid, and let M_α the dyadic fractional integral operator. Define:*

$$\beta(Q) = \frac{\sigma(Q)^{1/p'} w(Q)^{1/q}\, |Q|^{\alpha/n}}{|Q|}.$$

Set $[\sigma, w] := \sup_{Q\in\mathcal{D}} \beta(Q)$, then

$$\lambda^q w(\{I_\alpha f > \lambda\}) \lesssim [\sigma, w]^q\, \|f\|_{L^p(\sigma)}^q. \tag{4.10}$$

Proof. Let $\mathcal{D}_\lambda$ be the maximal elements of $\mathcal{D}$ contained in Q_0 such that $|Q|^{\alpha/n}\langle f\sigma\rangle_Q >$

λ. Since $\langle f\sigma\rangle_Q = \langle f\rangle_Q^\sigma \langle\sigma\rangle_Q$, there holds:

$$\lambda^q w\{Mf > \lambda\} \leqslant \sum_{Q\in\mathcal{D}_\lambda} \lambda^q w(Q) \leqslant \sum_{Q\in\mathcal{D}_\lambda} |Q|^{\frac{q\alpha}{n}} \langle\sigma\rangle_Q^q w(Q) \left(\langle f\rangle_Q^\sigma\right)^q$$
$$\leqslant [\sigma, w]^q \sum_{Q\in\mathcal{D}_\lambda} \sigma(Q)^{\frac{q}{p}} \left(\langle f\rangle_Q^\sigma\right)^q.$$

Given the disjointness of the sets $Q \in \mathcal{D}_\lambda$, (4.10) is immediate for $p = 1$. For $p > 1$, notice the sequence $\{\sigma(Q)^{q/p}\}_{Q\in\mathcal{D}_\lambda}$ is p, q–Carleson with respect to the measure σ. □

4.4 A One-Bump Condition for the Fractional Maximal Operator

By Lemma 2.4, Theorem 4.3 follows from the following lemma. We remark that while the following proof does not make explicit use of the Sawyer Maximal testing inequalities in [84], the proof does use some of the same ideas.

Lemma 4.11. *Let $1 < p \leqslant q < \infty$, and let σ and w be two weights. Given a dyadic grid $\mathcal{D}$, let M_α be the dyadic fractional maximal operator. Let ϵ_q be a monotonic increasing function on $(1, \infty)$ that satisfies $\int_1^\infty \frac{dt}{t\epsilon_q^q(t)} = 1$. Define*

$$\beta(Q) := \frac{\sigma(Q)^{1/p'} w(Q)^{1/q}}{|Q|^{1-\alpha/n}} \rho_\sigma^{1/p}(Q)\epsilon_q(\rho_\sigma(Q)),$$

Set $\lceil\sigma, w\rceil := \sup_{Q\in\mathcal{Q}} \beta(Q)$, then

$$\|M_\alpha f\sigma\|_{L^q(w)} \lesssim \lceil\sigma, w\rceil \|f\|_{L^p(\sigma)}.$$

Proof. Let $\mathcal{S}$ be any sparse subset of $\mathcal{D}$. By Remark 2.5 we need to verify

$$\int_{Q_0} \left| \sum_{Q\in\mathcal{S}:Q\subset Q_0} |Q|^{\alpha/n} \langle f\sigma\rangle_Q 1_{E_Q}(x) \right|^q w(x)dx \lesssim \lceil\sigma, w\rceil^q \|f\|_{L^p(\sigma)}^q. \qquad (4.12)$$

Let $\mathcal{Q}_k := \{Q \in \mathcal{S}, Q \subset Q_0 : \lceil \sigma, w \rceil 2^{-k} \leqslant \beta(Q) \leqslant \lceil \sigma, w \rceil 2^{-k+1}\}$. We will show

$$\int_{Q_0} \left| \sum_{Q \in \mathcal{Q}_k} |Q|^{\alpha/n} \langle f\sigma \rangle_Q 1_{E_Q}(x) \right|^q w(x)dx \lesssim (2^{-k})^q \lceil \sigma, w \rceil^q \|f\|^q_{L^p(\sigma)}. \qquad (4.13)$$

Taking q^{th} roots and summing over k will imply (4.12).

Using the identity $\langle f\sigma \rangle_Q = \langle \sigma \rangle_Q \langle f \rangle^\sigma_Q$ and the pairwise disjointness of the sets E_Q, (4.13) will follow from:

$$\sum_{Q \in \mathcal{Q}_k} \frac{|Q|^{q\alpha/n} \sigma(Q)^q w(Q)}{|Q|^q} (\langle f \rangle^\sigma_Q)^q \lesssim (2^{-k})^q \lceil \sigma, w \rceil^q \|f\|^q_{L^p(\sigma)}.$$

Thus, by the Carleson Embedding Theorem (Theorem F), it is enough to verify:

$$\frac{1}{\sigma(P)^{q/p}} \sum_{Q \in \mathcal{Q}_k : Q \subset P} \frac{|Q|^{q\alpha/n} \sigma(Q)^q w(Q)}{|Q|^q} \lesssim (2^{-k})^q \lceil \sigma, w \rceil^q,$$

for all $P \in \mathcal{Q}_k$. Using the fact that $\beta(Q) \simeq 2^{-k} \lceil \sigma, w \rceil$ for $Q \in \mathcal{Q}_k$ we estimate:

$$\begin{aligned} \sum_{Q \in \mathcal{Q}_k : Q \subset P} \frac{|Q|^{q\alpha/n} \sigma(Q)^q w(Q)}{|Q|^q} &= \sum_{Q \in \mathcal{Q}_k : Q \subset P} \frac{|Q|^{q\alpha/n} \sigma(Q)^{q/p'} w(Q)}{|Q|^q} \sigma(Q)^{q/p} \\ &\simeq (2^{-k})^q \lceil \sigma, w \rceil^q \sum_{Q \in \mathcal{Q}_k : Q \subset P} \frac{\sigma(Q)^{q/p}}{\rho_\sigma(Q)^{q/p} \epsilon_q^q(\rho_\sigma(Q))}. \end{aligned}$$

We want to show that the sum above is dominated by $\sigma(P)^{q/p}$. To this end, set $\mathcal{S}_r = \{Q \in \mathcal{Q}_k, Q \subset P : 2^{r-1} \leqslant \rho_\sigma(Q) \leqslant 2^r\}$. Thus, the sum above is dominated by

$$\sum_{r=0}^\infty \frac{1}{2^{rq/p} \epsilon_q^q(2^r)} \sum_{Q \in \mathcal{S}_r} \sigma(Q)^{q/p}.$$

Appealing to the summability condition on ϵ_q, it suffices to show that

$$\sum_{Q \in \mathcal{S}_r} \sigma(Q)^{q/p} \leqslant 2^{qr/p} \sigma(P)^{q/p}.$$

Let $\mathcal{S}_r^*$ be the maximal elements in $\mathcal{S}_r$. Observe that for fixed $S^* \in \mathcal{S}_r^*$, and for any $P \subset S^*$, there holds:

$$\left(\int_{E_Q} \langle 1_{S^*\sigma}\rangle_Q 1_Q\right)^{q/p} \leqslant \left(\int_{E_Q} \sup_{P\in\mathcal{D}} \langle 1_{S^*}\sigma\rangle_P 1_P\right)^{q/p}.$$

Since the sets E_Q are pairwise disjoint, $|Q| \simeq |E_Q|$, and $\int_{S^*} \sup_{P\in\mathcal{D}}\langle 1_{S^*}\sigma\rangle_P \leqslant \sigma(S^*)\rho_\sigma(S^*) \simeq 2^r\sigma(S^*)$ for $S^* \in \mathcal{S}_r^*$, we estimate

$$\begin{aligned}
\sum_{Q\in\mathcal{S}_r} \sigma(Q)^{q/p} &\leqslant \sum_{S^*\in\mathcal{S}_r^*}\sum_{Q\subset S^*}\left(\int_{E_Q} \sup_{P\in\mathcal{D}}\langle 1_{S^*}\sigma\rangle_P 1_P\right)^{q/p} \\
&\leqslant \sum_{S^*\in\mathcal{S}_r^*}\left(\int_{S^*} \sup_{P\in\mathcal{D}}\langle 1_{S^*}\sigma\rangle_P 1_P\right)^{q/p} \\
&\lesssim 2^{qr/p}\sum_{S^*\in\mathcal{S}_r^*} \sigma^{q/p}(S^*).
\end{aligned}$$

Using the disjointness of the sets $S^* \in \mathcal{S}_r^*$, the sum in the last line above is dominated by $\sigma(P)^{q/p}$, completing the proof. □

4.5 A One-Bump Condition

By Lemma 2.4, Theorem 4.4 follows from the following lemma.

Lemma 4.14. *Let $1 < p \leqslant q < \infty$, and let σ and w be two weights. Given a dyadic grid $\mathcal{D}$, let $I_\alpha^{\mathcal{D}}$ be the dyadic fractional integral operator. Let ϵ_p be a monotone increasing function on $(1,\infty)$ such that $\int_1^\infty \frac{dt}{t\epsilon_p^p(t)} = 1$, and similarly for $\epsilon_{q'}$. Define*

$$\beta(Q) := \frac{\sigma(Q)^{1/p'}w(Q)^{1/q}\,|Q|^{\alpha/n}}{|Q|}\rho_\sigma(Q)^{1/p}\epsilon_p(\rho_\sigma(Q))\rho_w(Q)^{1/q'}\epsilon_{q'}(\rho_w(Q)).$$

Set $\lfloor\sigma,w\rfloor := \sup_{Q\in\mathcal{Q}}\beta(Q)$, then

$$\left\|I_\alpha^{\mathcal{D}}(f\sigma)\right\|_{L^q(w)} \lesssim \lfloor\sigma,w\rfloor\,\|f\|_{L^p(\sigma)}.$$

Proof. We proceed by duality. Let $f \in L^p(\sigma)$ and $g \in L^{q'}(w)$. We use the identity: $\langle f\sigma\rangle_Q = \langle f\rangle_Q^\sigma \langle\sigma\rangle_Q$, where $\langle f\rangle_Q^\sigma := \sigma(Q)^{-1}\int_Q f\sigma$. From the definition of $\lfloor\sigma, w\rfloor$,

$$\begin{aligned}
\left\langle \sum_{Q\in\mathcal{Q}} |Q|^{\alpha/n}\langle f\sigma\rangle_Q 1_Q, gw \right\rangle &= \sum_{Q\in\mathcal{Q}} \langle f\rangle_Q^\sigma \langle g\rangle_Q^w |Q|^{\alpha/n}\langle\sigma\rangle_Q w(Q)\,|Q|^{\alpha/n} \\
&= \sum_{Q\in\mathcal{Q}} \langle f\rangle_Q^\sigma \sigma(Q)^{\frac{1}{p}} \langle g\rangle_Q^w w(Q)^{\frac{1}{q'}} \frac{\sigma(Q)^{\frac{1}{p'}} w(Q)^{\frac{1}{q}} |Q|^{\alpha/n}}{|Q|^{1-\frac{\alpha}{n}}} \\
&\lesssim \lfloor\sigma, w\rfloor \sum_{Q\in\mathcal{Q}} \frac{\langle f\rangle_Q^\sigma \sigma(Q)^{\frac{1}{p}}}{\rho_\sigma^{\frac{1}{p}}(Q)\epsilon_p(\rho_\sigma(Q))} \frac{\langle g\rangle_Q^w w(Q)^{\frac{1}{q'}}}{\rho_w^{\frac{1}{q'}}(Q)\epsilon_{q'}(\rho_w(Q))}.
\end{aligned}$$

The inner product in the first line is in $L^2(dx)$. By Hölder's inequality, it suffices to show that

$$\left(\sum_{Q\in\mathcal{S}} \frac{\sigma(Q)}{\rho_\sigma(Q)\epsilon_p^p \rho_\sigma(Q)} (\langle f\rangle_Q^\sigma)^p\right)^{\frac{1}{p}} \quad \text{and} \quad \left(\sum_{Q\in\mathcal{S}} \frac{w(Q)^{p'/q'}}{\rho_w^{p'/q'}(Q)\epsilon_{q'}^{p'}\rho_w(Q)} (\langle g\rangle_Q^w)^{q'}\right)^{\frac{1}{p'}}$$

are dominated by $\|f\|_{L^p(\sigma)}$ and $\|g\|_{L^{q'}(w)}$, respectively. Since $p \leqslant q$, it follows that $q' \leqslant p'$, so by the the Carleson Embedding Theorem (Theorem F), it suffices to show the following hold for all $Q_0 \in \mathcal{S}$:

(1)
$$\sum_{Q\in\mathcal{S}:Q\subset P} \frac{\sigma(Q)}{\rho_\sigma(Q)\epsilon_p^p(\rho_\sigma(Q))} \lesssim \sigma(Q_0)$$

(2)
$$\sum_{Q\in\mathcal{S}:S\subset P} \frac{w(Q)^{p'/q'}}{\rho_w^{p'/q'}(Q)\epsilon_{q'}^{p'}(\rho_w(Q))} \langle g\rangle_Q^{wq'} \lesssim w^{p'/q'}(Q_0).$$

But we omit the details since the proofs are similar to those in Lemma 4.11. $\square$

4.6 A Separated Bump Condition

From Remark 2.5 and Lemma A, it is enough to show

$$\int_{Q_0} \left| \sum_{Q \in \mathcal{Q}: Q \subset Q_0} |Q|^{\alpha/n} \langle \sigma \rangle_Q 1_Q(x) \right|^q w(x) dx \lesssim [[\sigma, w]]_{\alpha,p,q} \sigma(Q_0)^{q/p}$$

for any sparse collection $\mathcal{Q}$ and $Q_0 \in \mathcal{Q}$ (the dual testing condition follows identically). For the remainder, fix a root Q_0 and let $\mathcal{Q}$ be a sparse collection of cubes contained in Q_0. Fix α, p, q in the respective appropriate range; we'll ignore these fixed indices where there is no confusion. It remains to show

$$\left\| \sum_{Q \in \mathcal{Q}} |Q|^{\alpha/n} \langle \sigma \rangle_Q 1_Q \right\|_{L^q(w, Q_0)} \lesssim [[\sigma, w]]^{1/q} \sigma(P)^{1/p}.$$

For $Q \in \mathcal{Q}$, define

$$\beta(Q) := \left(|Q|^{\alpha/n} \langle \sigma \rangle_Q \right)^{q/p'} \langle w \rangle_Q \rho_\sigma(Q) \varepsilon_q \left(\rho_\sigma(Q) \right).$$

For integers a and r, set $\mathcal{Q}^{a,r} := \{Q \in \mathcal{Q} : \beta(Q) \simeq 2^a, \rho(Q) \simeq 2^r\}$; notice $\mathcal{Q}^{a,r}$ is empty for a large enough. Construct a stopping family $\mathcal{S}$ for the σ fractional averages: let $\mathcal{S}$ be the minimal subset of $\mathcal{Q}^{a,r}$ containing the maximal cubes in $\mathcal{Q}^{a,r}$ such that whenever $S \in \mathcal{S}$, the maximal cubes $Q \subset S$, $Q \in \mathcal{Q}^{a,r}$ with $|Q|^{\alpha/n} \langle \sigma \rangle_Q > 4|S|^{\alpha/n} \langle \sigma \rangle_S$ are also in $\mathcal{S}$. Denote by $Q^{\mathcal{S}}$ the $\mathcal{S}$–parent of Q. Partition $\mathcal{Q}^{a,r}$ into $\mathcal{Q}^{a,r}_k$, those cubes in $\mathcal{Q}^{a,r}$ such that $|Q|^{\alpha/n} \langle \sigma \rangle_Q \simeq 2^{-k} |Q^{\mathcal{S}}|^{\alpha/n} \langle \sigma \rangle_{Q^{\mathcal{S}}}$. We temporarily denote $\mathcal{Q}^{a,r}_k$ by $\mathcal{Q}'$. We will show

$$\left\| \sum_{Q \in \mathcal{Q}'} |Q|^{\alpha/n} \langle \sigma \rangle_Q 1_Q \right\|_{L^q(w)} \lesssim 2^{-k} \left[\sum_{S \in \mathcal{S}} |S|^{q\alpha/n} \langle \sigma \rangle_S^q w(S) \right]^{1/q}, \qquad (4.15)$$

where summing over $k \geqslant -2$ gives

$$\left\| \sum_{Q \in \mathcal{Q}^{a,r}} |Q|^{\alpha/n} \langle \sigma \rangle_Q 1_Q \right\|_{L^q(w)} \lesssim \left[\sum_{S \in \mathcal{S}} |S|^{q\alpha/n} \langle \sigma \rangle_S^q w(S) \right]^{1/q}. \tag{4.16}$$

Define for each $S \in \mathcal{S}$

$$\Phi_S := \sum_{Q \in \mathcal{Q}' : Q^{\mathcal{S}} = S} |Q|^{\alpha/n} \langle \sigma \rangle 1_Q \qquad \text{and} \qquad \Phi_{S,\ell} := \Phi_S 1_{\{\Phi_S \simeq \ell 2^{-k} |S|^{\alpha/n} \langle \sigma \rangle_S\}}.$$

Since $\sum\limits_{S \in \mathcal{S}} \Phi_{S,\ell}$ is geometric for fixed $\ell \in \mathbb{Z}^+$, Hölder's inequality yields some

$$\left(\sum_{\ell \geqslant 1} \sum_{S \in \mathcal{S}} \Phi_{S,\ell} \right)^q \lesssim \sum_{\ell \geqslant 1} \ell^{2q/q'} \left(\sum_{S \in \mathcal{S}} \Phi_{S,\ell} \right)^q \simeq \sum_{\ell \geqslant 1} \ell^{2q/q'} \sum_{S \in \mathcal{S}} \Phi_{S,\ell}^q. \tag{4.17}$$

It is apparent that we need the following distributional estimate.

Lemma 4.18. *There holds*

$$w \left\{ \Phi_S > \lambda 2^{-k} |S|^{\alpha/n} \langle \sigma \rangle_S \right\} \lesssim 2^{-\lambda} w(S).$$

Proof. The inequality is immediate in the case w is Lebesgue measure from sparseness of $\mathcal{Q}$. Notice that we have for $Q \in \mathcal{Q}'$ with $Q^{\mathcal{S}} = S$,

$$\langle w \rangle_Q \simeq \frac{2^a}{2^r \varepsilon_q(2^r)} (2^{-k} \langle \sigma \rangle_S |S|^{\alpha/n})^{-q/p'} =: \tau_S,$$

where the equivalence is independent of S. Denote by $\mathcal{Q}^*$ the maximal cubes in $\mathcal{Q}'$. Since the $\left\{ \Phi_S > \lambda 2^{-k} |S|^{\alpha/n} \langle \sigma \rangle_S \right\}$ is the union of the maximal cubes $P \in \mathcal{Q}'$ with

$P^{\mathcal{S}} = S$ and $\inf_{x\in P} \Phi_S(x) > \lambda 2^{-k}|S|^{\alpha/n}\langle\sigma\rangle_S$, hence a disjoint union, it follows that

$$\begin{aligned} w\left\{\Phi_S > \lambda 2^{-k}|S|^{\alpha/n}\langle\sigma\rangle_S\right\} &\simeq \tau_S \left|\left\{\Phi_S > \lambda 2^{-k}|S|^{\alpha/n}\langle\sigma\rangle_S\right\}\right| \\ &\lesssim \tau_S \left(2^{-(\lambda-1)} \sum_{Q^*\in\mathcal{Q}^*} |Q^*|\right) \\ &\simeq 2^{-\lambda} \sum_{Q^*\in\mathcal{Q}^*} w(Q^*). \end{aligned}$$

The collection $\mathcal{Q}^*$ is disjoint, so the proof is complete. □

Since $\{\Phi_{S,\ell} > \lambda 2^{-k}|S|^{\alpha/n}\langle\sigma\rangle_S\}$ is constant for $0 < \lambda < \frac{\ell}{2}$ and is empty for $\lambda > \ell$, we have

$$\begin{aligned} \int_{Q_0} \Phi_{S,\ell}^q dw &= 2^{-kq}|S|^{q\alpha/n}\langle\sigma\rangle_S^q \int_0^\infty q\lambda^{q-1} w\{\Phi_{S,\ell} > \lambda 2^{-k}|S|^{\alpha/n}\langle\sigma\rangle_S\} d\lambda \\ &\lesssim 2^{-kq}|S|^{q\alpha/n}\langle\sigma\rangle_S^q \left[\left(\frac{\ell}{2}\right)^q 2^{-\ell/2} w(S) + \frac{\ell}{2} q\ell^{q-1} 2^{-\ell/2} w(S)\right] \\ &\simeq 2^{-kq}|S|^{q\alpha/n}\langle\sigma\rangle_S^q \left[\ell^q 2^{-\ell/2} w(S)\right], \end{aligned}$$

where the second inequality is the application of Lemma 4.18. Recalling (4.17), this gives (4.15).

For each S define E_S to be S less the members of $\mathcal{S}$ properly contained in S. Let $\mathcal{S}^*$ be the maximal elements of $\mathcal{S}$. Since $\beta(S) \simeq 2^a$ and $\rho(S) \simeq 2^r$ for all $S \in \mathcal{S}$, the right hand side of (4.16) is equivalent to

$$\begin{aligned} \frac{2^a}{2^r \varepsilon_q(2^r)} \sum_{S\in\mathcal{S}} \left(|S|^{\alpha/n}\langle\sigma\rangle_S\right)^{q/p} |S| &\lesssim \frac{2^a}{2^r \varepsilon_q(2^r)} \left(\sum_{S^*\in\mathcal{S}^*} \sum_{S^*\supseteq S\in\mathcal{S}} \int_{E_S} M_\alpha(1_{S^*}\sigma)^{\frac{q}{p}} dx\right) \\ &\simeq \frac{2^a}{\varepsilon_q(2^r)} \left(\sum_{S^*\in\mathcal{S}^*} \sigma(S^*)^{q/p}\right) \\ &\lesssim \left[(2^{1/q})^a \frac{1}{\varepsilon_q^{1/q}(2^r)} \sigma(Q_0)^{1/p}\right]^q. \end{aligned}$$

The first inequality above follows from $|S| \simeq |E_S| = \int_{E_S} dx$, and the third by comparing q^{th} roots and remembering $p \leqslant q$. Take q^{th} roots above to attain the desired inequality. Summing the last quantity over integers $r \geqslant 0$ evokes the integrability condition on ε_q; summing over relevant integers a completes the proof.

CHAPTER 5

COMMUTATORS WITH FRACTIONAL INTEGRAL OPERATORS

Recall the Calderón-Zygmund operators:

$$Tf(x) := \int_{\mathbb{R}^n} K(x,y)f(y)\,dy, \quad x \notin \operatorname{supp} f,$$

where the kernel satisfies the standard size and smoothness estimates:

$$|K(x,y)| \leqslant \frac{C}{|x-y|^n},$$
$$|K(x+h,y) - K(x,y)| + |K(x,y+h) - K(x,y)| \leqslant C\frac{|h|^\delta}{|x-y|^{n+\delta}},$$

for all $|x-y| > 2|h| > 0$ and a fixed $\delta \in (0,1]$.

To contrast the Calderón-Zygmund operators with the fractional integral operators, note for example that fractional integral operators are positive, which in many cases makes them easier to work with. On the other hand, the fractional integral operators do not commute with dilations and therefore can never boundedly map $L^p(dx)$ to itself. Additionally, the kernel of the fractional integral operator does not satisfy the standard estimates above. Therefore, the theory of fractional integral operators is not just a subset of the theory of Calderón–Zygmund operators. Because of this, results which are known for Calderón-Zygmund operators also need to be proved for the fractional integral operators.

In this chapter we will characterize the triples (b,μ,λ), where b is a function and μ and λ are $A_{p,q}$ weights (to be defined shortly), such that the commutator $[b, I_\alpha]$ is bounded from $L^p(\mu^p)$ to $L^q(\lambda^q)$. Commutators with fractional integral operators were first studied in [12].

Our characterization will be in terms of the norm of b in a certain weighted BMO space, built from the weights μ and λ. This is an adaptation to the fractional integral setting of a viewpoint introduced by Bloom [5] in 1985, and recently investigated by the first Holmes, Lacey and Wick in [33, 32]. Specifically, Bloom characterized $\|[b, H] : L^p(\mu) \to L^p(\lambda)\|$, where H is the Hilbert transform and μ, λ are A_p weights, in terms of $\|b\|_{BMO(\nu)}$, where $BMO(\nu)$ is the weighted BMO space associated with the weight $\nu := \mu^{1/p}\lambda^{-1/p}$. Recall that the Hilbert transform is the one-dimensional prototype for Calderón-Zygmund operators, a role played by the fractional integral operators in $\mathbb{R}^n$.

A modern dyadic proof of Bloom's result was recently given in [33], and the techniques developed were then used to extend the result to all Calderón-Zygmund operators in [32]. In particular, it was proved that

$$\|[b, T] : L^p(\mu) \to L^p(\lambda)\| \leqslant c\|b\|_{BMO(\nu)}, \tag{5.1}$$

for all A_p weights μ, λ, and all Calderón-Zygmund operators T on $\mathbb{R}^n$, for some constant c depending on n, T, μ, λ and p. Specializing to the fractional integral operators, a lower bound was also proved. The center of the proof of (5.1) is the Hytönen Representation Theorem, which allows one to recover T from averaging over some dyadic operators, called dyadic shifts. Then the upper bound reduced to these dyadic operators.

We take a similar approach here, where the role of the dyadic shifts will be played by the dyadic version of the fractional integral operator I_α, given by:

$$I_\alpha^{\mathcal{D}} f := \sum_{Q \in \mathcal{D}} |Q|^{\alpha/n} \langle f \rangle_Q 1_Q. \tag{5.2}$$

Our main result is:

Theorem 5.3. *Suppose that* $\alpha/n + 1/q = 1/p$ *and* $\mu, \lambda \in A_{p,q}$*. Let* $\nu := \mu\lambda^{-1}$*. Then:*

$$\|[b, I_\alpha] : L^p(\mu^p) \to L^q(\lambda^q)\| \simeq \|b\|_{BMO(\nu)}.$$

It is important to observe that we require that each weight belong to a certain $A_{p,q}$ class and this will imply that $\mu\lambda^{-1}$ is an A_2 weight and in particular, an A_∞ weight. Standard properties of these weight classes will be used throughout the chapter, with out tracking dependencies on the particular weight characteristics. The liberal use of these properties indicates the subtleties involved in the general two–weight setting. For an excellent account of this and other topics related to fractional integral operators, see [18].

The chapter is organized as follows. In Section 5.1, we will give the requisite background material and definitions. Note, however, that most of the material not relating strictly to fractional integral operators (such as the Haar system, A_p weights, and weighted BMO) is standard and was also needed in [32] where it is discussed in more detail. In Section 5.2 we will briefly discuss how the fractional integral operator can be recovered as an average of dyadic operators. In Section 5.3 we will prove $\|[b, I_\alpha] : L^p(\mu^p) \to L^q(\lambda^q)\| \lesssim \|b\|_{BMO(\nu)}$ and in Section 5.4, we will prove the reverse inequality: $\|b\|_{BMO(\nu)} \lesssim \|[b, I_\alpha] : L^p(\mu^p) \to L^q(\lambda^q)\|$.

5.1 Background and Notation

5.1.1 The Haar System

Let $\mathcal{D}$ be a dyadic grid on $\mathbb{R}^n$ and let $Q \in \mathcal{D}$. For every $\varepsilon \in \{0,1\}^n$, let h_Q^ε be the usual Haar function defined on Q. For convenience, we write $\varepsilon = 1$ if $\varepsilon = (1, 1, \ldots, 1)$. Note that, in this case, $\int h_Q^1 = 1$. Otherwise, if $\varepsilon \neq 1$, then $\int h_Q^\varepsilon = 0$. Moreover, recall that $\{h_Q^\varepsilon\}_{Q\in\mathcal{D},\varepsilon\neq 1}$ forms an orthonormal basis for $L^2(\mathbb{R}^n)$. For a function f, a cube $Q \in \mathcal{D}$ and $\epsilon \neq 1$, we denote

$$\widehat{f}(Q, \epsilon) := \langle f, h_Q^\epsilon \rangle,$$

where $\langle \cdot, \cdot \rangle$ is the usual inner product in $L^2(\mathbb{R}^n)$.

5.1.2 A_p Classes and Weighted BMO

Let w be a weight on $\mathbb{R}^n$, that is, a locally integrable, almost everywhere positive function. For a subset $Q \subset \mathbb{R}^n$ we denote

$$w(Q) := \int_Q w \, dx \quad \text{and} \quad \langle w \rangle_Q := \frac{w(Q)}{|Q|}.$$

Given $1 < p < \infty$, a weight w is said to belong to the Muckenhoupt A_p class provided that:

$$[w]_{A_p} := \sup_Q \langle w \rangle_Q \langle w^{1-p'} \rangle_Q^{p-1} < \infty,$$

where p' denotes the Hölder conjugate of p, and the supremum is over all cubes $Q \subset \mathbb{R}^n$. Moreover, $w \in A_p$ if and only if $w^{1-p'} \in A_{p'}$ and, in this case, $[w^{1-p'}]_{A_{p'}} = [w]_{A_p}^{p'-1}$. Furthermore, if $1 < p < q < \infty$, then $A_p \subset A_q$, with $[w]_{A_q} \leqslant [w]_{A_p}$ for all $w \in A_p$.

For a dyadic lattice $\mathcal{D}$, recall the dyadic square function:

$$(S_{\mathcal{D}} f)^2 = \sum_{P \in \mathcal{D}, \varepsilon \neq 1} \left| \widehat{f}(Q, \epsilon) \right|^2 \frac{1_Q}{|Q|}.$$

Another property of A_p weights which will be useful for us is the following well–known weighted Littlewood–Paley Theorem:

Theorem 5.4. *Let $w \in A_p$, then:*

$$\|S_{\mathcal{D}} : L^p(w) \to L^p(w)\| \simeq c(n, p, [w]_{A_p}).$$

For a weight w on $\mathbb{R}^n$, the weighted BMO space BMO(w) is defined to be the space of all locally integrable functions b that satisfy:

$$\|b\|_{BMO(w)} := \sup_Q \frac{1}{w(Q)} \int_Q |b - \langle b \rangle_Q| \, dx < \infty,$$

where the supremum is over all cubes Q in $\mathbb{R}^n$. For a general weight, the definition of the BMO norm is highly dependent on its L^1 average. But, if the weight is A_∞, one is free to replace the L^1-norm by larger averages. Namely, for $w \in A_p$, define

$$\|b\|_{BMO^{p'}(w)} := \sup_Q \left(\frac{1}{w(Q)} \int_Q |b - \langle b \rangle_Q|^{p'} \, dw' \right)^{\frac{1}{p'}},$$

where w' denotes the conjugate weight $w^{1-p'}$. Then there holds

$$\|b\|_{BMO(w)} \leqslant \|b\|_{BMO^{p'}(w)} \leqslant C(n, p, [w]_{A_\infty}) \|b\|_{BMO(w)}. \tag{5.5}$$

The proof is similar to the proof in the unweighted case. In particular, the first inequality is a straightforward application of Hölder's inequality and the second inequality follows from a suitable John–Nirenberg property (which requires a suitable Calderón–Zygmund decomposition). The details are in [64].

For a dyadic grid $\mathcal{D}$ on $\mathbb{R}^n$, we define the dyadic versions of the norms above by taking supremum over $Q \in \mathcal{D}$ instead of over all cubes Q in $\mathbb{R}^n$, and denote these spaces by $BMO_{\mathcal{D}}(w)$ and $BMO^{p'}_{\mathcal{D}}(w)$. Clearly $BMO(w) \subset BMO_{\mathcal{D}}(w)$ for any choice of $\mathcal{D}$, and the equivalence in (5.5) also holds for the dyadic versions of these spaces.

A fact which will be crucial to our proof is the following:

Lemma 5.6. *If $w \in A_2$, then*

$$|\langle b, \Phi \rangle| \lesssim \|b\|_{BMO^2_{\mathcal{D}}(w)} \|S_{\mathcal{D}}\Phi\|_{L^1(w)}.$$

This comes from a duality relationship between dyadic weighted BMO spaces and dyadic weighted Hardy spaces. For a more detailed discussion and a proof of this fact, see [32, Section 2.6]. We remark here that Lemma 5.6 was also fundamental for the proof of the upper bound (5.1) in [32], essentially for the following reason: if μ, λ are A_p weights, then $\nu := \mu^{1/p}\lambda^{-1/p}$ is an A_2 weight. Thus the duality statement above applied to ν eventually yields, through Hölder's inequality, some bounds in terms of $L^p(\mu)$ and $L^{p'}(\lambda)$ norms. This is also the strategy we will adapt accordingly to the fractional integral case, which makes use of $A_{p,q}$ classes instead. We discuss these next.

5.1.3 $A_{p,q}$ Classes

Throughout this subsection, α, n, p, q are fixed and satisfy $1/p - 1/q = \alpha/n$. We recall first the fractional maximal operator,

$$M_\alpha f := \sup_Q |Q|^{\alpha/n} \langle |f| \rangle_Q 1_Q,$$

with the supremum being over all cubes Q. This was first introduced in [65], where it was used to prove weighted inequalities for I_α, a result analogous to the classic result [14] of Coifman and Fefferman, relating the Hardy-Littlewood maximal operator and singular integrals. We will be working with the dyadic version of this operator, $M_\alpha^{\mathcal{D}}$, defined for a dyadic grid $\mathcal{D}$ just as above, but only taking supremum over $Q \in \mathcal{D}$.

Also in [65] was introduced a generalization of A_p classes for the fractional integral setting: we say that a weight w belongs to the $A_{p,q}$ class provided that

$$[w]_{A_{p,q}} := \sup_Q \langle w^q \rangle_Q \langle w^{-p'} \rangle_Q^{q/p'} < \infty.$$

See [80, 77, 20, 19, 18] for other generalizations.

We will use the following important result concerning $A_{p,q}$ weights due to, for example, Sawyer and Muckenhoupt and Wheeden [83, 84, 65]:

Theorem 5.7. *Let w be a weight. Then the following are equivalent:*

(i) $w \in A_{p,q}$,

(ii) $\left\| M_\alpha^{\mathcal{D}} : L^p(w^p) \to L^q(w^q) \right\| \simeq C(n, \alpha, p, [w]_{A_{p,q}})$,

(iii) $\left\| I_\alpha^{\mathcal{D}} : L^p(w^p) \to L^q(w^q) \right\| \simeq C(n, \alpha, p, [w]_{A_{p,q}})$.

We now make two observations about $A_{p,q}$ weights which will be particularly useful to us. First, we note that:

$$\text{If } w \in A_{p,q}, \text{ then: } w^p \in A_p,\ \ w^{-p'} \in A_{p'},\ \ w^q \in A_q, \text{ and } \ w^{-q'} \in A_{q'}, \tag{5.8}$$

where all weights above have Muckenhoupt characteristics bounded by powers of $[w]_{A_{p,q}}$. To see that $w^p \in A_p$, first notice $w \in A_{p,q}$ if and only if $w^q \in A_{q_0}$, with $[w^q]_{A_{q_0}} = [w]_{A_{p,q}}$, where

$$q_0 := 1 + q/p' = q(1 - \alpha/n).$$

Since the A_p classes are increasing and $q_0 < q$, we have that $w^q \in A_q$. In turn, this gives that $w^{-q'} = (w^q)^{1-q'} \in A_{q'}$. The other two statements in (5.8) follow in a similar fashion from the fact that $w \in A_{p,q}$ if and only if $w^{-1} \in A_{q',p'}$.

Second, suppose that $\mu, \lambda \in A_{p,q}$ and let $\nu := \mu\lambda^{-1}$. Since $\mu^p, \lambda^p \in A_p$, Hölder's inequality implies $\nu \in A_2$ (with $[\nu]_{A_2}^p \leqslant [\mu^p]_{A_p}[\lambda^p]_{A_p}$), a fact which will be used in proving the upper bound. Moreover, we claim that for any cube Q:

$$\mu^p(Q)^{1/p}\lambda^{-q'}(Q)^{1/q'} \lesssim \nu(Q)|Q|^{\alpha/n}, \tag{5.9}$$

a fact which will be useful in proving the lower bound. To see this, note first that

$$\langle \mu^p \rangle_Q^{1/p} \langle \mu^{-p'} \rangle_Q^{1/p'} \lesssim 1 \quad \text{and} \quad \langle \lambda^{-q'} \rangle_Q^{1/q'} \langle \lambda^q \rangle_Q^{1/q} \lesssim 1,$$

which simply come from $\mu^p \in A_p$ and $\lambda^q \in A_q$. Since $p' > q'$, Hölder implies

$$\begin{aligned} \left(\frac{1}{|Q|} \int_Q \mu^{-q'}\, dx \right)^{1/q'} &\leqslant \left(\frac{1}{|Q|} \left(\int_Q \mu^{-p'}\, dx \right)^{q'/p'} \left(\int_Q \, dx \right)^{1-q'/p'} \right)^{1/q'} \\ &= \left(\frac{1}{|Q|} \int_Q \mu^{-p'}\, dx \right)^{1/p'}, \end{aligned}$$

and hence $\langle \mu^{-q'} \rangle_Q^{1/q'} \leqslant \langle \mu^{-p'} \rangle_Q^{1/p'}$. Combining these estimates gives:

$$\langle \mu^p \rangle_Q^{1/p} \langle \lambda^{-q'} \rangle_Q^{1/q'} \lesssim \frac{1}{\langle \mu^{-p'} \rangle_Q^{1/p'}} \frac{1}{\langle \lambda^q \rangle_Q^{1/q}} \lesssim \frac{1}{\langle \mu^{-q'} \rangle_Q^{1/q'} \langle \lambda^q \rangle_Q^{1/q}} \leqslant \frac{1}{\langle \nu^{-1} \rangle_Q} \leqslant \langle \nu \rangle_Q.$$

The last two inequalities are more application of Hölder's inequality and the fact that $\nu^{-1} = \mu^{-1}\lambda$. This proves (5.9).

5.2 *Averaging Over Dyadic Fractional Integral Operators*

In this section, we show that I_α can be recovered from (5.2) by averaging over dyadic lattices. The proof here is modified (and abridged) from the proof in [74], but it is possible to modify any of the proofs in, for example, [75, 35, 48]. For the sake of clarity, we only give the proof for the one–dimensional case.

Given an interval $[a,b)$ (it is not too important that the interval be closed on the left and open on the right) of length r, we can create a dyadic lattice, $\mathcal{D}_{a,r}$ in a standard way. In particular, $\mathcal{D}_{a,r}$ is the dyadic lattice on $\mathbb{R}$ with intervals of length $r2^{-k}$, $k \in \mathbb{Z}$, and the point a is not in the interior of any of the intervals in $\mathcal{D}_{a,r}$. For example, $\mathcal{D}_{0,1}$ is the standard dyadic lattice on $\mathbb{R}$. For a given lattice $\mathcal{D}_{a,r}$, we let $\mathcal{D}^k_{a,r}$ denote the intervals in $\mathcal{D}_{a,r}$ with length $r2^{-k}$. In this section we slightly abuse notation and let $h^1_I = |I|^{-1/2}\, 1_I$.

Define:

$$\mathbb{P}^0_{(a,r)} f(x) := \sum_{I \in \mathcal{D}^0_{a,r}} |I|^{\alpha} \left\langle f, h^1_I \right\rangle h^1_I(x).$$

With r and x fixed, we can parameterize the dyadic grids by the set $(-r,0]$ and we can give this set the probability measure da/r. For a fixed $x \in \mathbb{R}$, we want to compute:

$$\mathbb{E}(\mathbb{P}^0_{(a,r)} f(x)) = \int_{-r}^{0} \mathbb{P}^0_{(a,r)} f(x) \frac{da}{r}.$$

Let $\tau_t f(x) := f(x+t)$ be the translation operator and note that $\mathbb{P}_{a-t}\tau_t = \tau_t \mathbb{P}_a$. From this it easily follows that $\mathbb{E}\mathbb{P}^0_{(a,r)}\tau_t = \tau_t \mathbb{P}^0_{(a,r)}$. That is, $\mathbb{E}\mathbb{P}^0_{(a,r)}$ is given by convolution. Let:

$$\mathbb{E}\mathbb{P}^0_{(a,r)} f(x) = F_{0,r} * f(x).$$

We want to compute $F_{0,r}$. First, note that $\mathbb{P}^0_{a,r}$ is convolution with the function

$\frac{r^\alpha}{r}1_{[-r/2,r/2]}$. Therefore, we have:

$$\begin{aligned} F_{0,r} * f(x) &= \mathbb{E}\mathbb{P}^0_{(a,r)} f(x) \\ &= \mathbb{E}\mathbb{P}^0_{(a/2,r)} f(x) \int_{x-r/2}^{x+r/2} \int_{\mathbb{R}} f(s) \frac{r^\alpha}{r} 1_{-r/2,r/2}(t-s)\, ds \frac{dt}{r}. \end{aligned}$$

Using Fubini, we see that:

$$F_{0,r}(x) = \int_{x-r/2}^{x+r/2} \frac{r^\alpha}{r} 1_{[-r/2,r/2]}(t) \frac{dt}{r} = \frac{r^\alpha}{r} 1_{[-r/2,r/2]}(x) \left(1 - \left|\frac{x}{r}\right|\right) = \frac{r^\alpha}{r} F_{0,1}(x/r).$$

Now, fix an $r \in [1,2)$ and define:

$$F_r = \sum_{n \in \mathbb{Z}} F_{0,2^n r}.$$

The grids $\mathcal{D}^k_{a,r}, k \in \mathbb{Z}$ can be unioned to form a dyadic lattice (here a is fixed). Call r the calibre of the dyadic lattice. Convolution with F_r is averaging over all the dyadic lattices $\mathcal{D}_{a,r}$ with fixed calibre r. That is:

$$F_r * f = \mathbb{E}\mathbb{P}_{\mathcal{D}_{a,r}} f.$$

Finally, we need to average over $r \in [1,2)$. Set $\int_1^2 F_r(x) \frac{dr}{r} := F(x)$ and compute:

$$\begin{aligned} F(x) &= \int_1^2 F_r(x) \frac{dr}{r} \\ &= \int_1^2 \sum_{n \in \mathbb{Z}} F_{0,2^n r}(x) \frac{dr}{r} \\ &= \int_0^\infty F_{0,\rho}(x) \frac{d\rho}{\rho} \\ &= \int_0^\infty F_{0,1}\left(\frac{x}{\rho}\right) \frac{\rho^\alpha}{\rho^2} d\rho \\ &= \int_0^\infty 1_{-1/2,1/2}\left(\frac{x}{\rho}\right)\left(1 - \left|\frac{x}{\rho}\right|\right) \frac{\rho^\alpha}{\rho^2} dr. \end{aligned}$$

Now, if $x > 0$, making the change of variable $t = x/\rho$, we see:

$$F(x) = \frac{x^\alpha}{x}\int_0^\infty F_{0,1}(y)\frac{dy}{y^\alpha} = c_\alpha \frac{1}{x^{1-\alpha}}.$$

A similar computation for $x < 0$ yields $F(x) = c_\alpha \frac{1}{|x|^{1-\alpha}}$.

5.3 The Weighted Inequality

The decomposition in Section 5.2 means that the upper bound in Theorem 5.3 follows from the following, where the implied constants are independent of the dyadic lattice:

Lemma 5.10. *Suppose that* $\alpha/n + 1/q = 1/p$ *and* $\mu, \lambda \in A_{p,q}$. *Let* $\nu := \mu\lambda^{-1}$. *Then:*

$$\left\| [b, I_\alpha^{\mathcal{D}}] : L^p(\mu^p) \to L^q(\lambda^q) \right\| \lesssim \|b\|_{BMO(\nu)}.$$

Proof. We show that $[b, I_\alpha^{\mathcal{D}}]$ can be decomposed as the sum of four operators which will be fairly easy to bound. First note that for $\varepsilon \neq 1$, there holds:

$$I_\alpha^{\mathcal{D}} h_Q^\varepsilon = \sum_{P\in\mathcal{D}:P\subsetneq Q} |P|^{\alpha/n} h_Q^\varepsilon(P) 1_P = \Bigg(\sum_{P\in\mathcal{D}:P\subsetneq Q} |P|^{\alpha/n} 1_P\Bigg) h_Q^\varepsilon = c_\alpha |Q|^{\alpha/n} h_Q^\varepsilon.$$

Similarly,

$$I_\alpha^{\mathcal{D}} 1_Q = (1 + c_\alpha) |Q|^{\alpha/n} 1_Q + |Q| \sum_{R\in\mathcal{D}:Q\subsetneq R} |R|^{\alpha/n} \frac{1_R}{|R|}.$$

Using these computations:

$$I_\alpha^{\mathcal{D}}(h_P^\epsilon h_Q^\eta) = \begin{cases} c_\alpha |P\cap Q|^{\frac{\alpha}{n}} h_P^\epsilon h_Q^\eta & \text{, if } P \neq Q \text{ or if } P = Q \text{ and } \epsilon \neq \eta; \\ (1 + c_\alpha)|Q|^{\frac{\alpha}{n}} \frac{1_Q}{|Q|} + \sum_{R\supsetneq Q} |R|^{\frac{\alpha}{n}} \frac{1_R}{|R|} & \text{, if } P = Q \text{ and } \epsilon = \eta. \end{cases}$$

Thus:

$$[h_P^\epsilon, I_\alpha^{\mathcal{D}}]h_Q^\eta = \begin{cases} c_\alpha h_Q^\eta(P) h_P^\epsilon \left(|Q|^{\frac{\alpha}{n}} - |P|^{\frac{\alpha}{n}}\right) & \text{, if } P \subsetneq Q; \\ -|Q|^{\frac{\alpha}{n}} \frac{1_Q}{|Q|} - \sum_{R \supsetneq Q} |R|^{\frac{\alpha}{n}} \frac{1_R}{|R|} & \text{, if } P = Q \text{ and } \epsilon = \eta; \\ 0 & \text{, if } Q \subsetneq P \text{, or if } Q = P \text{ and } \epsilon \neq \eta. \end{cases}$$

Expressing b and f in terms of their Haar coefficients, we obtain that

$$[b, I_\alpha^{\mathcal{D}}]f = \sum_{P,Q \in \mathcal{D}} \sum_{\epsilon,\eta \neq 1} \widehat{b}(P,\epsilon)\widehat{f}(Q,\eta)[h_P^\epsilon, I_\alpha^{\mathcal{D}}]h_Q^\eta.$$

Using this, there holds

$$[b, I_\alpha^{\mathcal{D}}]f = c_\alpha T_1 f - c_\alpha \Pi_{b,\alpha}^{(0,1,0)} f - \Pi_{b,\alpha}^{(0,0,1)} f - T_2 f,$$

where:

$$\begin{aligned} \Pi_{b,\alpha}^{(0,1,0)} f &:= \sum_{Q \in \mathcal{D}, \epsilon \neq 1} \widehat{b}(Q,\epsilon) \langle f \rangle_Q |Q|^{\frac{\alpha}{n}} h_Q^\epsilon; \\ \Pi_{b,\alpha}^{(0,0,1)} f &:= \sum_{Q \in \mathcal{D}, \epsilon \neq 1} \widehat{b}(Q,\epsilon) \widehat{f}(Q,\epsilon) |Q|^{\frac{\alpha}{n}} \frac{1_Q}{|Q|}; \\ T_1 f &:= \sum_{P \in \mathcal{D}, \epsilon \neq 1} \widehat{b}(P,\epsilon) \left(\sum_{Q \supsetneq P, \eta \neq 1} \widehat{f}(Q,\eta) h_Q^\eta(P) |Q|^{\frac{\alpha}{n}} \right) h_P^\epsilon; \\ T_2 f &:= \sum_{P \in \mathcal{D}, \epsilon \neq 1} \widehat{b}(P,\epsilon) \widehat{f}(P,\epsilon) \left(\sum_{Q \supsetneq P} |Q|^{\frac{\alpha}{n}} \frac{1_Q}{|Q|} \right). \end{aligned}$$

We will show that all of these operators are bounded $L^p(\mu^p) \to L^q(\lambda^q)$. Below, all implied constants are allowed to depend on $n, \alpha, p, [\mu]_{A_{p,q}}$, and $[\lambda]_{A_{p,q}}$. Also all inner products below are taken with respect to dx and therefore it is enough to show:

$$|\langle Tf, g \rangle| \lesssim \|b\|_{BMO(\nu)} \|f\|_{L^p(\mu^p)} \|g\|_{L^{q'}(\lambda^{-q'})},$$

for each of the four operators above (this is because the dual of $L^q(\lambda^q)$ with respect to the unweighted inner product is $L^{q'}(\lambda^{-q'})$). The idea, which is taken from [33, 32], is to write the bilinear form, $\langle Tf, g\rangle$ as $\langle b, \Phi\rangle$ and then show that $\|S_{\mathcal{D}}\Phi\|_{L^1(\nu)}$ is controlled by $\|f\|_{L^p(\mu^p)}\|g\|_{L^{q'}(\lambda^{-q'})}$; by the weighted $H^1 - \text{BMO}$ duality, this is enough to prove the claim.

The estimates for the two paraproducts are almost identical, and we only give the proof for $\Pi_{b,\alpha}^{(0,1,0)}$. First with

$$\Phi := \sum_{Q\in\mathcal{D},\epsilon\neq 1} \langle f\rangle_Q |Q|^{\frac{\alpha}{n}} \widehat{g}(Q,\epsilon) h_Q^\epsilon,$$

there holds:

$$\left\langle \Pi_{b,\alpha}^{(0,1,0)} f, g\right\rangle = \langle b, \Phi\rangle\,.$$

Then:

$$(S_{\mathcal{D}}\Phi)^2 = \sum_{Q\in\mathcal{D},\epsilon\neq 1} |\langle f\rangle_Q|^2 |Q|^{\frac{2\alpha}{n}} |\widehat{g}(Q,\epsilon)|^2 \frac{1_Q}{|Q|} \leqslant (M_\alpha f)^2 (S_{\mathcal{D}} g)^2.$$

Therefore,

$$\|S_{\mathcal{D}}\Phi\|_{L^1(\nu)} \leqslant \|M_\alpha f\|_{L^q(\mu^q)} \|S_{\mathcal{D}} g\|_{L^{q'}(\lambda^{-q'})} \lesssim \|f\|_{L^p(\mu^p)} \|g\|_{L^{q'}(\lambda^{-q'})},$$

where the last inequality follows from Theorem 5.7 for the fractional maximal function, and from Theorem 5.4 and the fact that $\lambda^{-q'} \in A_{q'}$ for the dyadic square function. The proof for $\Pi_{b,\alpha}^{(0,0,1)}$ is very similar, and we omit the details.

Now let us look at T_1. As above, we have $\langle T_1 f, g\rangle = \langle b, \Phi\rangle$, with

$$\Phi := \sum_{P\in\mathcal{D},\epsilon\neq 1} \widehat{g}(P,\epsilon) \left(\sum_{Q\supsetneq P,\eta\neq 1} \widehat{f}(Q,\eta) h_Q^\eta(P) |Q|^{\frac{\alpha}{n}} \right) h_P^\epsilon,$$

Then:

$$(S_{\mathcal{D}}\Phi)^2 \leqslant \sum_{P\in\mathcal{D},\epsilon\neq 1} |\widehat{g}(P,\epsilon)|^2 \left(\sum_{Q\supsetneq P,\eta\neq 1} \langle |f| \rangle_Q |Q|^{\frac{\alpha}{n}} \right)^2 \frac{1_P}{|P|}$$
$$\leqslant (I_\alpha^{\mathcal{D}}|f|)^2 (S_{\mathcal{D}} g)^2.$$

From Theorem 5.7 and Theorem 5.4, it follows that

$$\|S_{\mathcal{D}}\Phi\|_{L^1(\nu)} \leqslant \|I_\alpha^{\mathcal{D}}|f|\|_{L^q(\mu^q)} \|S_{\mathcal{D}} g\|_{L^{q'}(\lambda^{-q'})} \lesssim \|f\|_{L^p(\mu^p)} \|g\|_{L^{q'}(\lambda^{-q'})}.$$

The estimates for T_2 are similar and we omit the details. □

5.4 The Reverse Weighted Inequality

In this section, we prove the lower bound in Theorem 5.3, which follows immediately from the Lemma below. In particular, we will show the following:

Lemma 5.11. *For all cubes,* Q*:*

$$\frac{1}{\nu(Q)} \int_Q |b(x) - \langle b \rangle_Q| \, dx \lesssim \|[b, I_\alpha] : L^p(\mu^p) \to L^q(\lambda^q)\| \, .$$

Proof. The proof here follows along the lines of the proof in [11]. We first make some reductions. As with unweighted BMO, we can replace the $\langle b \rangle_Q$ with any constant. Indeed, there holds:

$$\frac{1}{\nu(Q)} \int_Q |b(x) - \langle b \rangle_Q| \, dx \leqslant \frac{1}{\nu(Q)} \int_Q |b(x) - C_Q| \, dx + \frac{|Q|}{\nu(Q)} |C_Q - \langle b \rangle_Q|$$
$$\leqslant \frac{2}{\nu(Q)} \int_Q |b(x) - C_Q| \, dx.$$

Second, let P be the cube with $l(P) = 4l(Q)$, where $l(Q)$ is the side length of Q, and with the same “bottom left corner” as Q. By the doubling property of A_∞ weights, there holds $\nu(P) \simeq \nu(Q)$, and therefore it is enough to prove:

$$\frac{1}{\nu(P)} \int_Q |b(x) - C_Q| \, dx \lesssim \|[b, I_\alpha] : L^p(\mu^p) \to L^q(\lambda^q)\| \,.$$

Finally, let P_R be the "upper right half" of P. Below, we will use $C_Q = \langle b \rangle_{P_R}$.

Now, for $x \in Q$ and $y \in P_R$ there holds:

$$\frac{|x-y|}{2\sqrt{n}\,|P|^{1/n}} \geqslant \frac{\sqrt{n}\,|Q|^{1/n}}{2\sqrt{n}\,|P|^{1/n}} = \frac{1}{8} \quad \text{and} \quad \frac{|x-y|}{2\sqrt{n}\,|P|^{1/n}} \leqslant \frac{\sqrt{n}\,|P|^{1/n}}{2\sqrt{n}\,|P|^{1/n}} \leqslant \frac{1}{2}.$$

The point is that there is a function, $K(x)$, that is smooth on $[-1, 1]^n$, has a smooth periodic extension to $\mathbb{R}^n$, and is equal to $|x|^{n-\alpha}$ for $1/8 \leqslant |x| \leqslant 1/2$. Therefore, for $x \in Q$ and $y \in P_R$ there holds:

$$\left(\frac{|x-y|}{2\sqrt{n}\,|P|^{1/n}}\right)^{n-\alpha} = K\left(\frac{x-y}{2\sqrt{n}\,|J|}\right).$$

Important for us is the fact that K has a Fourier expansion with summable coefficients.

We are now ready to prove the main estimate. First, let $\sigma(x) = \operatorname{sgn}(b(x) - \langle b \rangle_{P_R})$. Then:

$$\begin{aligned}
\int_Q |b(x) - \langle b \rangle_{P_R}| \, dx &= \frac{1}{|P_R|} \int_{\mathbb{R}} \int_{\mathbb{R}} (b(x) - b(y)) \sigma(x) 1_Q(x) 1_{P_R}(y) \, dy \, dx \\
&= \frac{1}{|P_R|} \int_{\mathbb{R}} \int_{\mathbb{R}} \frac{b(x) - b(y)}{\left(\frac{|x-y|}{2\sqrt{n}|P|}\right)^{n-\alpha}} \left(\frac{|x-y|}{2\sqrt{n}\,|P|}\right)^{n-\alpha} \sigma(x) 1_Q(x) 1_{P_R}(y) \, dy \, dx \\
&\simeq |P|^{-\alpha/n} \int_{\mathbb{R}} \int_{\mathbb{R}} \frac{b(x) - b(y)}{|x-y|^{n-\alpha}} K\left(\frac{x-y}{2\sqrt{n}\,|P|}\right) \sigma(x) 1_Q(x) 1_{P_R}(y) \, dy \, dx.
\end{aligned}$$

Observe that the integral above is positive, so the "$\simeq$" is not a problem. Expanding

K in its Fourier series:

$$K\left(\frac{x-y}{2\sqrt{n}\,|P|}\right) = \sum_k a_k e^{ikx/2\sqrt{n}|P|} e^{-iky/2\sqrt{n}|P|},$$

the integral above becomes:

$$\sum_k a_k \int_Q \int_{P_R} \frac{b(x)-b(y)}{|x-y|^{n-\alpha}} \sigma(x) e^{ikx/c|P| - iky/c|P|} dy dx = \sum_k a_k \int_{\mathbb{R}} h_k(x)[b, I_\alpha] f_k(x) dx,$$

where $h_k(x) = \sigma(x) e^{ikx/c|P|} 1_P(x)$ and $f_k(y) = e^{-iky/c|P|} 1_{P_R}(y)$. We control the integral by:

$$\begin{aligned}
\int_{\mathbb{R}} h_k(x)[b, I_\alpha] f_k(x) dx &\leqslant \|[b, I_\alpha] : L^p(\mu^p) \to L^q(\lambda^q)\| \, \|f_k\|_{L^p(\mu^p)} \|h_k\|_{L^{q'}(\lambda^{-q'})} \\
&= \|[b, I_\alpha] : L^p(\mu^p) \to L^q(\lambda^q)\| \, \mu^p(P_R)^{1/p} \lambda^{-q'}(P)^{1/q'} \\
&= \|[b, I_\alpha] : L^p(\mu^p) \to L^q(\lambda^q)\| \, \mu^p(P)^{1/p} \lambda^{-q'}(P)^{1/q'}.
\end{aligned}$$

By (5.9), this is dominated by:

$$\|[b, I_\alpha] : L^p(\mu^p) \to L^q(\lambda^q)\| \, |P|^\alpha \, \nu(P).$$

This completes the proof. $\square$

CHAPTER 6

OSCILLATORY AND RANDOM SINGULAR INTEGRALS

This chapter explores the theme of bounding singluar integral operators by sparse operators in the settings of (a) oscillatory singular integrals, and (b) discrete random operators. In both cases, we easily derive weighted inequalities. In the latter case, these are the first such weighted inequalities known. We state our results before providing a broader context.

Theorem 2.12 yields a non-trivial corollary:

Corollary 6.1. *For $1 < p < \infty$, the operator T_P, where $P = P(y)$ is of degree d, is bounded on $L^p(w)$, where w is a Muckenhoupt weight $w \in A_p$.*

Weak-type and weighted estimates for oscillatory singular integrals have been studied in this and more general contexts by various authors, see for instance [27, 28, 30, 31, 82]. Y. Ding and H. Liu [27] were interested in $L^p(w)$ inequalities for more general operators T. The approach of these authors entails many complications.

The method of proof of Theorem 2.12 is very simple, so we suspect that stronger results are possible. For instance, this Conjecture would imply nearly sharp A_p bounds, for all $1 < p < 2$.

Conjecture 6.2. *For $1 < r < \infty$, the operator T_P, where $P = P(y)$ is of degree d, for each bounded compactly supported function f, there is a sparse operator $\Lambda_{1,r}$ so that*

$$|\langle T_P f, g\rangle| \lesssim \Lambda_{1,r}(f, g).$$

It seems likely that the weak type argument of Chanillo and Christ [13] would establish the Conjecture for $r = 2$. Also see [45].

We turn to weighted inequalities for *discrete random Hilbert transforms* acting on functions on $\ell^2(\mathbb{Z})$.

Corollary 6.3. *For any $0 < \alpha < 1$, almost surely, the following holds: For all $1 + \alpha < p < \frac{1+\alpha}{\alpha}$, and weights w so that*

$$w^{1+\alpha} \in A_{(1+\alpha)(p-1)+1}, \qquad w \in A_{1+\frac{1}{(1+\alpha)(p'-1)}}, \tag{6.4}$$

we have $\|H_\alpha : \ell^p(w) \mapsto \ell^p(w)\| < \infty$. The implied constant only depends upon $[w^{1+\alpha}]_{A_{(1+\alpha)(p-1)+1}}$, and $[w]_{A_{1+\frac{1}{\alpha(p'-1)}}}$. The same inequality holds for M_α.

The study of these questions was initiated by Bourgain [8], as an elementary example of a sequence of integers for which one could derive ℓ^p inequalities, with the sequence of integers also having asymptotic density zero. Various aspects of these questions have been studied, both in ℓ^p and at the weak $(1,1)$ endpoints [81, 9, 63, 87, 53]. We are not aware of any result in the literature that proves a weighted estimate in this sort of discrete setting. (If the set of integers has full density, it is easy to transfer weighted estimates.)

There is a subtle difference between the Hilbert transform and the maximal function in this random setting. In particular, more should be true for the maximal function. Prompted by the work of LaVictoire [53], we pose

Conjecture 6.5. *For $0 < \alpha < 1/2$, almost surely, for all $1 < r < 2$, and finitely supported functions f, g, there is a sparse operator $\Lambda_{1,r}$ so that*

$$\langle M_\alpha f, g\rangle \lesssim \Lambda_{1,r}(f, g).$$

We turn to the context for our paper. The concept of sparse operators arose from

Lerner's remarkable median inequality [56]. It's application to weighted inequalities was advanced by several authors, with a high point of this development being Lerner's argument [57] showing that the weighted norm of Calderón-Zygmund operators is comparable to that of the norms of sparse operators. This lead to the question of pointwise control, namely Theorem **??**. First established by Conde-Alonso and Rey [16], also see Lerner and Nazarov [54], Lacey [47] established Theorem **??** with a stopping time argument. The latter argument was extended by Bernicot, Frey and Petermichl [3] to a setting where the operators are generated by semigroups, including examples outside the scope of classical Calderón-Zygmund theory. For closely related developments see [60, 41]. The sparse bounds for commutators [55, 25] are remarkably powerful. Edging beyond the Calderón-Zygmund context, Benau, Bernicot and Frey [2] have supplied sparse bounds for certain Bochner-Riesz multipliers.

Very recently, Culiuc, di Plinio and Ou [23] have established a sparse domination result in a setting far removed from the extensions above: The trilinear form associated to the bilinear Hilbert transform is dominated by a sparse form. This is a surprising result, as the bilinear Hilbert transform has all the difficult features of the Hilbert transform, with additional oscillatory and arithmetic-like aspects. Wile the point of this chapter is to understand how general a technique 'domination by sparse' could be, there are plenty of additional directions that one could think about.

For instance, the interest in the oscillatory singular integrals is driven in part by their application to singular integrals defined on nilpotent groups. Implications of the sparse bound in this setting are unexplored.

After applying the known sparse bounds for singular integrals, for the remaining parts of the operator, there is a very simple interpolation argument which you can use in the bilinear setting. The notable point about the proofs are that they are quite easy, and yet deliver striking applications.

6.1 The Sparse Bilinear Bound of Oscillatory Singular Integrals

We prove here Theorem 2.12. Our conclusion is invariant under dilations of the operator. Hence, we can proceed under the assumption that $\|P\| = \sum_\alpha |\lambda_\alpha| = 1$. We can also assume that the polynomial P has no linear term, as it can be absorbed into the function f. Under these assumptions we prove

Theorem 6.6. *Let P be a polynomial without linear terms, and $\|P\| = 1$. Then, for bounded compactly supported functions f, g and $1 < r < \infty$, there is a sparse form Λ_1 and a $\eta > 0$ so that*

$$|\langle T_P f, g\rangle| \lesssim \Lambda_1(f,g) + \sum_{Q \in \mathcal{D} \,:\, |Q| \geqslant 1} \langle f\rangle_{Q,r} \langle g\rangle_{Q,r} |Q|^{1-\eta} \tag{6.7}$$

It is easy to see that this implies Theorem 2.12, since the second term on the right is restricted to dyadic cubes of volume at least one, and there is a gain of $|Q|^{-\eta}$. Moreover, we will see that this Theorem implies the weighted result.

Let $e(\lambda) = e^{i\lambda}$ for $\lambda \in \mathbb{R}$. If the kernel K of T is supported on $2B = \{y \,:\, |y| \leqslant 2\}$, then we have

$$|e(P(y))K(y) - K(y)| \lesssim \mathbf{1}_{2B}(y)|y|^{-n+1},$$

so that $|T_P f - Tf| \lesssim Mf$. Both T and M admit pointwise domination by sparse forms, hence also by bilinear forms. (This is the main result of [47].)

Thus, we can proceed under the assumption that the kernel K is not supported on B. We can then write

$$K = \sum_{j=1}^{\infty} \varphi_j$$

where φ_j is supported on $2^{j-1}B \setminus 2^{j-2}B$, with $\|\nabla^s \varphi_j\|_\infty \lesssim 2^{-nj-sj}$, for $s = 0, 1$.

We use shifted dyadic grids, $\mathcal{D}_t$, for $1 \leqslant t \leqslant 3^n$. These grids have the property that

$$\{\tfrac{1}{3}Q \ : \ Q \in \mathcal{D}_t,\ \ell Q = 2^k, 1 \leqslant t \leqslant 3^n\}$$

form a partition of $\mathbb{R}^n$. Throughout, $\ell Q = |Q|^{1/n}$ is the side length of the cube Q. We fix a dyadic grid $\mathcal{D}_t$ throughout the remainder of the argument, and set $\mathcal{D}_+ = \{Q \ : \ \ell Q > 2^{10}\}$. Define

$$I_Q f = \int e(P(y))\varphi_k(y)(\mathbf{1}_{\frac{1}{3}Q} f)(x-y)\ dy, \qquad \ell Q = 2^{k+2}.$$

Note that $I_Q f$ is supported on Q, and that we have suppressed the dependence on P, which we will continue below.

The basic estimate is then this Lemma.

Lemma 6.8. *For each cube Q with $|Q| \geqslant 1$ and $1 < r < 2$, there holds*

$$|\langle I_Q f, g\rangle| \lesssim 2^{-\eta k}\langle f\rangle_{Q,r}\langle g\rangle_{Q,r}|Q|, \tag{6.9}$$

where $\eta = \eta(d,n,r) > 0$.

Theorem 6.6 follows immediately from this Lemma. The oscillatory nature of the problem exhibits itself in the next Lemma. Write

$$I_Q^* I_Q \phi(x) = \mathbf{1}_{\frac{1}{3}Q}(x) \cdot \int_{\frac{1}{3}Q} K_Q(x,y)\phi(y)\ dy.$$

Lemma 6.10. *For each cube $Q \in \mathcal{D}_+$, and $x \in \frac{1}{3}Q$, we have*

$$|K_Q(x,y)| \lesssim |Q|^{-1}\mathbf{1}_{Z_Q}(x-y) + |Q|^{-1-\epsilon}\mathbf{1}_Q(x)\mathbf{1}_Q(y),$$

where $Z_Q \subset Q$ has measure at most $(\ell Q)^{-\epsilon}|Q|$, where $\epsilon = \epsilon(n,d) > 0$.

This Lemma is well known, see for instance [85, Lemma 4.1]. Here is how we use the Lemma. Using Cauchy-Schwartz, we have

$$\begin{aligned}\|I_Q f\|_2^2 &\lesssim |Q|^{-1} \int_Q \int_{Z_Q} |f(x)||f(x-y)| \, dy dx + |Q|^{-\epsilon} \langle f \rangle_{Q,1}^2 |Q| \\ &\lesssim |Q|^{-\epsilon/n} \|f\mathbf{1}_Q\|_2^2.\end{aligned}$$

We also have the trivial but rarely used $\|I_Q f\|_\infty \lesssim |Q|^{-1}\|f\mathbf{1}_Q\|_1$. By Riesz Thorin interpolation, there holds with $\ell Q = 2^k$,

$$\|I_Q f\|_{r'} \lesssim 2^{-\eta k}|Q|^{-1+2/r'}\|f\mathbf{1}_Q\|_r, \qquad 1 < r \leqslant 2, \ r' = \tfrac{r}{r-1}.$$

Above, $\eta = \eta(\epsilon, r)$. But, this immediately implies (6.9). Namely,

$$\begin{aligned}|\langle I_Q f, g\rangle| &\lesssim \|I_Q f\|_{r'} \|g\mathbf{1}_Q\|_r \\ &\lesssim 2^{-\eta k}|Q|^{-1+2/r'}\|f\mathbf{1}_Q\|_r\|g\mathbf{1}_Q\|_r \\ &= 2^{-\eta k}\langle f\rangle_{Q,r}\langle g\rangle_{Q,r}|Q|.\end{aligned}$$

(Alternatively, one can just use bilinear interpolation.)

We now give the weighted result.

Proof of Corollary 6.1. The qualitative result that T_P is bounded on $L^p(w)$ for $w \in A_p$, $1 < p < \infty$ is as follows. Given $w \in A_p$, recall that the dual weight is $\sigma = w^{1-p'}$. Then, it is equivalent to show that

$$|\langle T_P(f\sigma), gw\rangle| \lesssim C_{[w]_{A_p}} \|f\|_{L^p(\sigma)} \|g\|_{L^{p'}(w)}.$$

Using the sparse domination from (6.7), we see that we need to prove the corresponding bound for the terms on the right in (6.7). Now, it is well known [57] that

$$\Lambda_1(f,g) \lesssim [w]_{A_p}^{\max\{1, \frac{1}{p-1}\}} \|f\|_{L^p(w)} \|g\|_{L^{p'}(w)}.$$

Indeed, this is a key part of the proof of the A_2 Theorem by sparse operators.

So, it remains to consider the second term on the right in (6.7). For each $k \in \mathbb{N}$, we have by Proposition 6.15, $k \in \mathbb{Z}$,

$$\sum_{Q \in \mathcal{D}\,:\,|Q| = 2^{nk}} \langle f \rangle_{Q,r} \langle g \rangle_{Q,r} |Q| \lesssim [w]_{A_p}^{1/p} [w]_{RH_r} [\sigma]_{RH_r} \|f\|_{L^p(w)} \|g\|_{L^{p'}(w)}.$$

As we recall in § 6.3, there is a $r = r([w]_{A_p}) > 1$ so that $[w]_{RH_r}[\sigma]_{RH_r} < 4$. And so the proof of the Corollary is complete.

Indeed, it is easy enough to make this step quantitative. For $2 < p < \infty$, the choice of r can be taken to satisfy $r - 1 > c[w]_{A_p}^{-1}$, which then means that the choice of $\eta = \eta(r)$ in (6.7) is at least as big is $c[w]_{A_p}^{-1}$. Then, our bound is

$$\langle T_P(\sigma f), gw \rangle \lesssim [w]_{A_p}^{1+\frac{1}{p}} \|f\|_{L^p(\sigma)} \|g\|_{L^{p'}(w)}, \qquad 2 < p < \infty.$$

We have no reason to believe that this estimate is sharp. □

6.2 Random Hilbert Transforms

The discrete Hilbert transform

$$Hf(x) = \sum_{n \neq 0} \frac{f(x-n)}{n}$$

satisfies a sparse bound: For all finitely supported functions f and g, there is a sparse operator Λ so that

$$|\langle Hf, g \rangle| \lesssim \Lambda_{1,1}(f, g). \tag{6.11}$$

This is a consequence of the main results of Theorem **??**. Recall the definition of H_α in Subsection 2.3.2; there is also stated there Theorem 2.13, which we prove here.

Notice that $\mathbb{E}H_\alpha f = Hf$, so it remains to consider the difference

$$
\begin{aligned}
H_\alpha f(x) - Hf(x) &:= \sum_{k=1}^{\infty} \sum_{n\,:\,2^{k-1}\leqslant |n| < 2^k} \frac{X_n - n^{-\alpha}}{n^{1-\alpha}} f(x-n) \\
&:= \sum_{k=1}^{\infty} T_k f(x).
\end{aligned}
$$

Above, we have passed directly to the distinct scales of the operator. We will subsequently write $Y_n = X_n - n^{-\alpha}$, which are independent mean zero random variables.

The crux of the matter are these two estimates:

Lemma 6.12. *Almost surely, for all* $0 < \epsilon < 1$, *and for all integers* k, *and* f, g *supported on an interval* I *of length* 2^k, *we have*

$$
|\langle T_k f, g\rangle| \lesssim \begin{cases} 2^{-k\frac{1-\alpha}{2}+\epsilon} \langle f\rangle_{I,2} \langle g\rangle_{I,2} |I| \\ 2^{k\alpha} \langle f\rangle_{I,1} \langle g\rangle_{I,1} |I| \end{cases} .
$$

The implied constant is random, but independent of $k \in \mathbb{N}$ *and the choice of functions* f, g.

Proof. The second bound follows trivially from $|Y_n|/n^{1-\alpha}\mathbf{1}_{2^{k-1}\leqslant|n|<2^k} \lesssim 2^{k(\alpha-1)}$. For the first bound, we clearly have

$$
|\langle T_k f, g\rangle| \leqslant \|T_k \ :\ \ell^2 \to \ell^2\| \cdot \langle f\rangle_{I,2} \langle g\rangle_{I,2} |I|,
$$

so it suffices to estimate the operator norm above. The assertion is that with high probability, the operator norm is small:

$$
\mathbb{P}\big(\|T_k \ :\ \ell^2 \to \ell^2\| > C\sqrt{k} 2^{-k\frac{1-\alpha}{2}}\big) \lesssim 2^{-k},
$$

provided C is sufficiently large. Combine this with the Borel Cantelli Lemma to prove the Lemma as stated.

By Plancherel's Theorem, the operator norm is equal to $\|Z(\theta)\|_{L^\infty(d\theta)}$, where

$$Z(\theta) := \sum_{n\,:\,2^k \leqslant |n| < 2^{k+1}} Y_n \frac{e^{2\pi i\theta}}{n^{1-\alpha}}.$$

The expression above is a random Fourier series, with frequencies at most 2^{k+2}. By Bernstein's Theorem for trigonometric polynomials, the $L^\infty(d\theta)$ norm can be estimated by testing the norm on at most 2^{k+3} equally spaced points in $\mathbb{T}$, that is, we have

$$\mathbb{P}\big(\|Z(\theta)\|_\infty > C\sqrt{k}2^{-k\frac{1-\alpha}{2}}\big) \lesssim 2^k \sup_\theta \mathbb{P}\big(|Z(\theta)| > C\sqrt{k}2^{-k\frac{1-\alpha}{2}}\big),$$

where we have simply used the union bound.

Now, $Z(\theta)$ is the sum of independent, mean zero random variables, which are bounded by one, and have standard deviation bounded by $c2^{-k\frac{1-\alpha}{2}}$. So by, for instance, the Bernstein inequality, it follows that

$$\mathbb{P}(|Z(\theta)| > C\sqrt{k}2^{-k\frac{1-\alpha}{2}}) \lesssim 2^{-2k},$$

for appropriate C. This completes the proof. □

From the previous Lemma, we have the Corollary below. It with the sparse bound for the Hilbert transform (6.11) completes the proof of Theorem 2.13, for the random Hilbert transform. The case for maximal averages is entirely similar.

Corollary 6.13. *Almost surely, for* $1+\alpha < r < 2$, *there is a* $\eta > 0$ *so that for all integers* k, *and all functions* f, g *supported on an interval* I *of length* 2^k, *we have*

$$|\langle T_k f, g\rangle| \lesssim 2^{-\eta k}\langle f\rangle_{I,r}\langle g\rangle_{I,r}|I|. \tag{6.14}$$

Proof. This follows from Lemma 6.12 by interpolation. The relevant interpolation parameter θ_0 at which we have only an epsilon loss in the interpolated estimate is given by

$$(1-\theta_0)\alpha = \theta_0 \frac{1-\alpha}{2},$$
$$\text{so} \quad \frac{1}{r_0} = \frac{1-\theta_0}{1} + \frac{\theta_0}{2}.$$

We see that $r_0 = 1+\alpha$. And so we conclude that for $r_0 = 1+\alpha < r < 2$, we have the required gain in the interpolated bound, which proves the Corollary. □

We now turn to the weighted inequalities of Corollary 6.3.

Proof of Corollary 6.3. For the deterministic Hilbert transform, we have the sharp bound of Petermichl [73], namely

$$\|H : \ell^p(w) \mapsto \ell^p(w)\| \lesssim [w]_{A_p}^{\max\{1, \frac{1}{p-1}\}}.$$

So, it remains to bound the terms in (6.14). By Proposition 6.15, we then need to see that the hypotheses on w, namely (6.4), imply that for some choice of $r > 1+\alpha$, we have

$$w \in A_p, \quad w \in RH_r, \quad \sigma = w^{1-p'} \in RH_r.$$

Recall that $v \in A_q \cap RH_s$ if and only if $v^s \in A_{s(q-1)+1}$. Now, by assumption, $w^{1+\alpha} \in A_{(1+\alpha)(p-1)+1}$. So, there is a $t > 1$ so that $w^{t(1+\alpha)} \in A_{(1+\alpha)(p-1)+1}$, and the A_q classes increase in q, so we conclude that $w \in A_p \cap RH_r$, for a $r > 1+\alpha$.

The second hypothesis is $w \in A_{1+\frac{1}{(1+\alpha)(p'-1)}}$. This is equivalent to

$$(w^{(1-p')})^{1+\alpha} \in A_{(1+\alpha)(p'-1)+1}.$$

Now, $w^{1-p'} = \sigma$ is the dual weight. So by the argument in the previous paragraph,

$\sigma \in RH_r$, for some $r > 1+\alpha$. So the proof is complete. $\square$

6.3 Weighted Inequalities

Let us recall the weighted estimates that we need for our corollaries. A function $w > 0$ is a *Muckenhoupt* A_p *weight* if

$$[w]_{A_p} = \sup_Q \Bigl[\frac{w^{\frac{1}{1-p}}(Q)}{|Q|}\Bigr]^{p-1} \frac{w(Q)}{|Q|} < \infty.$$

Above, we are conflating w as a measure and a density, thus $w^{\frac{1}{1-p}}(Q) = \int_Q w(x)^{\frac{1}{1-p}}\, dx$. We have these estimates, which are sharp in the A_p characteristic. They are an element of the sparse proof of the A_2 conjecture. (See [57] for a proof.)

Theorem G. *These estimates hold for all* $1 < p < \infty$*.*

$$\|\Lambda_{1,1} \ : \ L^p(w) \mapsto L^p(w)\| \lesssim [w]_{A_p}^{\max\{1,\frac{1}{p-1}\}}.$$

For our applications, we have a second class of operators, a simplified form of those introduced by Benau-Bernicot-Petermichl [2]. For our purposes, we need a much simplified version of their result. Define an additional characteristic of a weight, namely the *reverse Hölder* property.

$$[w]_{RH_r} = \sup_Q \frac{\langle w\rangle_{Q,r}}{\langle w\rangle_Q}.$$

Proposition 6.15. *Fix an integer* k*, and* $1 < r < 2$*. We have the bound below for all* $w \in A_p$*, where* $r \leqslant p \leqslant r' = \frac{r}{r-1}$*.*

$$\sum_{Q\in\mathcal{D}\,:\,|Q|=2^{nk}} \langle f\rangle_{Q,r}\langle g\rangle_{Q,r}|Q| \lesssim [w]_{A_p}^{1/p}[w]_{RH_r}[\sigma]_{RH_r}\|f\|_{L^p(w)}\|g\|_{L^{p'}(w)}$$

where $\sigma = w^{1-p'}$ *is the 'dual' weight to* w.

Let us recall these well known facts.

1. We always have $[w]_{A_p}, [w]_{RH_r} \geqslant 1$.
2. For $w \in A_p$ and $\sigma = w^{1-p'}$, the weight σ is locally finite, its 'dual' weight is w, and $[\sigma]_{A_{p'}} = [w]_{A_p}^{p'-1}$.
3. For every $w \in A_p$ there is a $r = r([w]_{A_p}) > 1$ so that $w \in RH_r$. (In particular, we can take r so that $r - 1 \simeq [w]_{A_p}^{-1}$, by [39]*Thm 2.3.)
4. For every $w \in A_p$, there is a $r = r([w]_{A_p}) > 1$ so that $w^r \in A_p$.
5. We have $w \in A_p \cap RH_r$ if and only if $w^r \in A_{r(p-1)+1}$, by [44].

Proof of Proposition 6.15. This inequality is rephrased in the self-dual way, namely setting $\sigma = w^{1-p'}$, it is equivalent to show that for $k \in \mathbb{Z}$,

$$\sum_{\substack{Q \in \mathcal{D} \\ |Q| = 2^{nk}}} \langle f\sigma \rangle_{Q,r} \langle gw \rangle_{Q,r} |Q| \lesssim [w]_{A_p}^{\frac{1}{p}} [\sigma]_{RH_r} [w]_{RH_r} \|f\|_{L^p(\sigma)} \|g\|_{L^{p'}(w)}. \tag{6.16}$$

Fix the integer k. We can assume that for $|Q| = 2^{nk}$, if f is not zero on Q, then $f\mathbf{1}_{3Q \setminus Q} \equiv 0$, and we assume the same for g. Then, set

$$f' = \sum_{Q \in \mathcal{D} \,:\, |Q| = 2^{nk}} \mathbf{1}_Q \left[\frac{1}{\sigma(Q)} \int_Q |f|^r \, d\sigma \right]^{1/r}$$

and likewise for g'. It is immediate that $\|f'\|_{L^p(\sigma)} \lesssim \|f\|_{L^p(\sigma)}$, thus in (6.16), it suffices to assume that $f = f'$. Then, we can even assume that f and g are supported on a single cube Q, and take the value 1 on that cube.

Then, write

$$\begin{aligned}\langle \sigma \mathbf{1}_Q \rangle_{Q,r} \langle w \mathbf{1}_Q \rangle_{Q,r} |Q| &\leqslant [\sigma]_{RH_r} [w]_{RH_r} \langle \sigma \mathbf{1}_Q \rangle_{Q,1} \langle w \mathbf{1}_Q \rangle_{Q,1} |Q| \\ &\leqslant [\sigma]_{RH_r} [w]_{RH_r} \langle \sigma \mathbf{1}_Q \rangle_{Q,1}^{1/p'} \langle w \mathbf{1}_Q \rangle_{Q,1}^{1/p} \cdot \sigma(Q)^{1/p} w(Q)^{1/p'} \\ &\leqslant [\sigma]_{RH_r} [w]_{RH_r} [w]_{A_p}^{1/p} \sigma(Q)^{1/p} w(Q)^{1/p'}.\end{aligned}$$

This is the inequality claimed.

□

Part II

A Learning Theory Approach to Compressive Sensing

CHAPTER 7
COMPRESSIVE SENSING

Compressed sensing is a modern data processing scheme that is proving useful in many scientific areas, such as MR imaging, radar, astronomy: see [1, 6, 62] for more details. The overarching goal is to reconstruct a *signal* $x \in \mathbb{R}^n$ from the *measurements* $Ax \in \mathbb{R}^m$ ($m \ll n$) given the sensing matrix $A \in \mathbb{R}^{m \times n}$ and some constraint on the set of signals. Without such a constraint, this is an ill-posed inverse problem, while more information about the signal x may make the objective approachable.

One common situation is that the signal is sparse: for a signal $x = (x^1, \ldots, x^n)$, we say x is s-sparse if $\left|\{x^j \neq 0\}\right| \leqslant s$. A successful program for reconstructing sparse signals is ℓ_1-minimization. This convex optimization algorithm is tractable and perfectly reconstructs s-sparse vectors (and well approximates them in the presence of noise) if the sensing matrix A has the (s, δ)-RIP with small enough δ [10]. A matrix A is said to have the (s, δ)-RIP if

$$(1-\delta)\|x-y\|_2^2 \leqslant \|Ax - Ay\|_2^2 \leqslant (1+\delta)\|x-y\|_2^2$$

for all pairs x, y of s-sparse vectors.

The focus of this Part is the analogue, i.e., dimension reducing quasi-isometric embeddings of sparse vectors, in the one-bit sensing framework. The purpose of this chapter is two-fold. First, it is natural to first introduce the somewhat easier case of linear measurements. After all, it was linear compressive sensing that came first, begetting one-bit sensing only as the intricacies of quantization in applications became apparent. Secondly, it turns out that some of the arguments in Chapter 8 extend naturally to the linear case. As such, we are able to prove a known results, Corllary 7.3, with a new and efficient proof. These results are organized in the following section

7.1 Corollaries

For s-sparse $x \in \mathbb{R}^n$, set $H_{x,\alpha} = \{y \in \mathbb{R}^n : \langle x, y \rangle > \alpha\}$, the half-space associated to x at height α. Denote by $H^{n,s}$ the set of such skew half-spaces in $\mathbb{R}^n$ associated to s-sparse signals: $H^{n,s} = \{H_{x,\alpha} : x \in \mathbb{R}^n, |\{x^j \neq 0\}| \leqslant s, \alpha \in \mathbb{R}\}$. The first result listed here gives a useful upper bound on the VC-dimension of $H^{n,s}$. VC-dimension is defined in Section 8.2

Corollary 7.1. $\mathrm{VC}(H^{n,s}) \lesssim s \log(n/s)$.

Proof. The analogue of the lower bound in Lemma 8.11 is achieved by shattering $B = \{0, e_1, \dots, e_s\}$, the standard basis vectors together with the origin. Any subset of $\{e_1, \dots, e_s\}$ can be achieved in the same was as in the Lemma, and an appropriate choice of α includes or excludes the origin as needed. The upper bound is Radon's theorem. This establishes $\mathrm{VC}(H^{s,s}) = s + 1$. The extension to $\mathrm{VC}(H^{n,s})$ is the same as the remainder of Section 8.2.2. □

Consider a sensing matrix with rows $\{g_k\}$ drawn from the standard Gaussian distribution. Then $\frac{1}{\sqrt{m}}A$ has the (s, δ)-RIP if and only if

$$\frac{1}{m}\left|\sum_{k=1}^{m} \langle g_k, z\rangle^2 - 1\right| < \delta$$

for all $2s$-sparse unit signals $z \in \mathbb{R}^n$. It is clear that $\{\{\langle \cdot, z\rangle^2 > \alpha\} : |\{z^j \neq 0\}| \leqslant 2s, \alpha \in \mathbb{R}\}$ is a subset of the set of all unions of pairs of skew half-spaces in $H^{n,2s}$. We have the following Lemma to control the VC-dimension of the latter.

Lemma 7.2. *If* $\mathrm{VC}(\mathcal{C}) = d$*, then* $\mathrm{VC}(\mathcal{C}^*) \leqslant 10d$*, where* $\mathcal{C}^* := \{B \cup C : B, C \in \mathcal{C}\}$.

Proof. Suppose $|X| = k$ and $\mathcal{C}^*$ shatters X. From Theorem H, $m^{\mathcal{C}}(k) \leqslant \left(\frac{ek}{d}\right)^d$. Hence $2^k \leqslant \left(\frac{ek}{d}\right)^{2d}$, or equivalently $k \leqslant 2d \log_2\left(\frac{ek}{d}\right)$. □

This Lemma and the discussion preceding it establishes the bound

$$\mathrm{VC}\left(\{\{\langle\cdot, z\rangle^2 > \alpha\} : |\{z^j \neq 0\}| \leqslant 2s,\ \alpha \in \mathbb{R}\}\right) \lesssim s \log(n/s).$$

An argument identical to the one in Section 8.3 proving Theorem 8.3 yields the following corollary. This is one of the fundamental results inspiring randomly drawn sensing matrices in compressive sensing. It is the smallest known portion of measurements necessary for the Restricted Isometry Property, and many practitioners and theorists believe that it cannot be beaten.

Corollary 7.3. *Let* $A \in \mathbb{R}^{m \times n}$ *with rows drawn independently from the standard Gaussian distribution. Then for any* $0 < \varepsilon, \delta < 1$ *and* $1 \leqslant s < n$, $\frac{1}{\sqrt{m}}A$ *has the* (s, δ)*-RIP with probability at least* $1 - \varepsilon$ *provided*

$$m \gtrsim \delta^{-2}\left[\log(2/\varepsilon) + s\log(n/s)\right].$$

CHAPTER 8

ONE-BIT SENSING

We study the dimension reducing *sign-linear* maps of one-bit compressed sensing. Associated to each $A \in \mathbb{R}^{m\times n}$ is the sign-linear map

$$\Phi_A : \mathbb{S}^{n-1} \to \mathcal{H}^m$$
$$\Phi_A x = \operatorname{sgn}(Ax),$$

where $\mathcal{H}^m$ is the Hamming Cube $\{\pm 1\}^m$, the *sgn* map is applied component-wise, and

$$\operatorname{sgn}(x) = \begin{cases} +1, & x > 0 \\ -1, & x \leqslant 0. \end{cases}$$

We restrict our attention to the sphere since any two signals that differ only in norm will have identical measurements. In the larger realm of compressed sensing, one-bit sensing is the case of extreme quantization: only the sign-bit of each linear measurement is preserved. The concept was initially suggested by Boufounos-Baraniuk [7] in 2008.

Let $\mathbb{S}^{n-1}_s$ denote the set of n-dimensional, unit length s-sparse signals. The (s,δ)-*Restricted Isometry Property*, or (s,δ)-RIP, analogue for Φ_A that we investigate is

$$\sup_{x,y\in\mathbb{S}^{n-1}_s} |d_{\mathcal{H}^m}(\Phi_A x, \Phi_A y) - d(x,y)| \leqslant \delta,$$

where $d(\cdot,\cdot)$ is geodesic distance on the sphere, and $d_{\mathcal{H}^m}(\cdot,\cdot)$ is the Hamming metric:

$$d_{\mathcal{H}^m}(a,b) := \tfrac{1}{m}\left|\{1 \leqslant k \leqslant m : a_k \neq b_k\}\right|.$$

The reader may notice that the one-bit RIP given above is single-scale, while the original RIP is multiscale. This modification is unavoidable; given $A \in \mathbb{R}^{m\times n}$ and $\varepsilon > 0$, there are $x, y \in \mathbb{S}^{n-1}_s$ such that $d(x,y) \leqslant \varepsilon$ and $d_{\mathcal{H}^m}(\Phi_A x, \Phi_A y) \geqslant \frac{1}{m}$. This formulation of the RIP has been studied theoretically, see [4, 43, 76]; it also plays a role in sparse signal recovery from one-bit measurements, e.g. [42, 43].

The effects of noise on a one-bit embedding is a natural concern, and we consider the case of additive white noise prior to quantization. When our sensing matrix is drawn randomly, we always assume the noise and matrix are independent. Associated to a matrix $A \in \mathbb{R}^{m\times n}$ and random vector $\eta \in \mathbb{R}^m$ is a one-bit embedding of the form

$$\Phi^\eta_A : \mathbb{R}^n \to \mathcal{H}^m$$
$$\Phi^\eta_A x = \operatorname{sgn}(Ax + \eta).$$

This is the the model of *systematic* noise, where the noise is randomly drawn but constant relative to the signals. In many applications, however, this is not the case, and the noise varies from signal to signal. In attempt to model such noise, we consider the one-bit embeddings associated to a matrix $A \in \mathbb{R}^{m\times n}$ and collection of random vectors $\{\eta(x)\} \subset \mathbb{R}^m$:

$$\Psi^\eta_A : \mathbb{R}^n \to \mathcal{H}^m$$
$$\Psi^\eta_A x = \operatorname{sgn}(Ax + \eta(x)).$$

The affects of the additive white noise on the RIP are analyzed by increasing the Gaussian measurements' dimension and lifting the sphere to a higher dimension by padding with σ^2 (and zero, depending on the noise model).

8.1 Outline and Main Results

For $x \in \mathbb{S}^{n-1}$, set $H_x = \{p \in \mathbb{S}^{n-1} : \langle p, x\rangle > 0\}$, the hemisphere associated to x. Denote by $H^{n,s}$ the set of hemispheres of $\mathbb{S}^{n-1}$ associated to s-sparse signals: $H^{n,s} = \{H_x : x \in \mathbb{S}^{n-1}_s\}$. The first result listed here gives a useful upper bound on the VC-dimension, defined in Section 8.2.2, of $H^{n,s}$. The result easily applies to half-spaces (Corollary 7.1), a well studied classification scheme in learning theory; it is well known that the VC-dimension of half-spaces in $\mathbb{R}^n$ indexed by s-sparse vectors is $\mathcal{O}(s \log n)$. The theorem below is slightly better (at least in some case, when n/s is small enough), but the author is unsure if it is known. We include the proof in Section 8.2.2 for completeness, and note that it is quite surprising to find the popular $s\log(n/s)$ quantity. Throughout, $x \lesssim y$ means there is an absolute $C > 0$ such that $x \leqslant Cy$.

Theorem 8.1. $\mathrm{VC}(H^{n,s}) \lesssim s\log(n/s)$.

Definition 8.2. Let $\Phi : \mathbb{S}^{n-1}_s \to \mathcal{H}^m$. We say Φ has the (s,δ)-RIP if

$$\sup_{x,y\in\mathbb{S}^{n-1}_s} |d_{\mathcal{H}^m}(\Phi x, \Phi y) - d(x,y)| \leqslant \delta.$$

Of note in Definition 8.2 is the metric $d(\cdot,\cdot)$, which is **not** the euclidean distance, but rather the geodesic distance on the sphere, normalized so that antipodal points are unit distance apart:

$$d(x,y) := \tfrac{1}{\pi}\arccos\langle x, y\rangle.$$

This choice of metric is natural since it is the expectation of $d_{\mathcal{H}^m}(\Phi_A x, \Phi_A y)$.

In Section 8.3 we employ a standard entropy integral argument to bound a supremum, indexed by pairs of s-sparse vectors. This is an alternative proof of a recent result of Bilyk-Lacey, the case of sparse vectors in [4, Theorem 1.14], which is:

Theorem 8.3. *Let $A \in \mathbb{R}^{m \times n}$ with rows drawn independently from the standard Gaussian distribution. Then for any $0 < \varepsilon, \delta < 1$ and $1 \leqslant s < n$, Φ_A has the (s, δ)-RIP with probability at least $1 - \varepsilon$ provided*

$$m \gtrsim \delta^{-2} \left[\log(2/\varepsilon) + s \log(n/s)\right].$$

The next theorems, proved in Section 8.4, are the import of the chapter. We consider the one-bit sign-linear maps with additive white noise prior to quantization. A curious detail about the result is that the error due to noise is not naturally expressed in the distortion parameter, nor the number of measurements or probability of success, but rather in the metric on the sphere. That is, if the sphere is endowed with a certain "distorted" geodesic metrics (8.4) and (8.5), the noisy embeddings have the (s, δ)-RIP with the same order of measurements and probability of success as determined in Theorem 8.3. Before stating the theorems, we define:

$$d^{\sigma}(x, y) := \tfrac{1}{\pi} \arccos\left(\tfrac{\langle x, y\rangle + \sigma^2}{1 + \sigma^2}\right) \tag{8.4}$$

$$d_{\sigma}(x, y) := \tfrac{1}{\pi} \arccos\left(\tfrac{\langle x, y\rangle}{1 + \sigma^2}\right). \tag{8.5}$$

We also define the following noisy versions of the one-bit RIP:

Definition 8.6. Let $\Phi : \mathbb{S}^{n-1}_s \to \mathcal{H}^m$. We say Φ has the $(s, \delta)^{\sigma}$-RIP if

$$\sup_{x, y \in \mathbb{S}^{n-1}_s} |d_{\mathcal{H}^m}(\Phi x, \Phi y) - d^{\sigma}(x, y)| \leqslant \delta,$$

and it has the $(s, \delta)_{\sigma}$-RIP if

$$\sup_{x, y \in \mathbb{S}^{n-1}_s} |d_{\mathcal{H}^m}(\Phi x, \Phi y) - d_{\sigma}(x, y)| \leqslant \delta.$$

Theorem 8.7. *Let $A \in \mathbb{R}^{m \times n}$ with rows drawn independently from the standard Gaus-*

sian distribution and $\eta \sim \mathcal{N}(\boldsymbol{0}, \sigma^2 I_m)$. *Then for any* $0 < \varepsilon, \delta < 1$ *and* $1 \leqslant s < n$, Φ_A^η *has the* $(s, \delta)^\sigma$*-RIP with probability at least* $1 - \varepsilon$ *provided*

$$m \gtrsim \delta^{-2} \left[\log(2/\varepsilon) + s \log(n/s)\right].$$

Theorem 8.8. *Let* $A \in \mathbb{R}^{m \times n}$ *with rows drawn independently from the standard Gaussian distribution and* $\{\eta(x)\}$ *a collection of pairwise independent random vectors indexed over the sphere with each* $\eta(x) \sim \mathcal{N}(\boldsymbol{0}, \sigma^2 I_m)$. *Then for any* $0 < \varepsilon, \delta < 1$ *and* $1 \leqslant s < n$, Ψ_A^η *has the* $(s, \delta)_\sigma$*-RIP with probability at least* $1 - \varepsilon$ *provided*

$$m \gtrsim \delta^{-2} \left[\log(2/\varepsilon) + s \log(n/s)\right].$$

Remark. *It is a common goal in signal processing to "eliminate" the noise. That is, one wishes to take enough measurements so that the noise is practically negligible. Theorem 8.7 and Theorem 8.8 demonstrate that this possibility is controlled by the variance in the Gaussian noise model. The empirical process of interest approaches a distorted metric, which is a deterministic object that necessarily deviates from the geodesic metric when* $\sigma^2 > 0$. *However, the distorted distances are close to geodesic distance at small scales, and when* $\sigma^2 \ll 1$, *the metrics are close globally.*

We conclude with Section 8.4.3, comparing the geodesic distance with the metrics defined in (8.4) and (8.5). Crude upper bounds on their differences give lower bounds on the number of Gaussian measurements needed for a noisy embedding to have the RIP into the Hamming cube with the *geodesic* metric prescribed to the sphere. While this result is appealing for obvious reasons, Theorem 8.7 and Theorem 8.8 may be more useful in practice, allowing the reader to appeal to the fact that the two metrics are indeed *very* close at small scales.

Corollary 8.9. *Let* $A \in \mathbb{R}^{m \times n}$ *with rows drawn independently from the standard Gaus-*

sian distribution and $\eta \sim \mathcal{N}(\boldsymbol{0}, \sigma^2 I_m)$. *Then for any* $0 < \varepsilon < 1$ *and* $\delta > 1 - \frac{1}{\pi}\arccos\left(\frac{\sigma^2-1}{\sigma^2+1}\right)$, Φ_A^η *has the* (s, δ)*-RIP with probability at least* $1 - \varepsilon$ *provided*

$$m \gtrsim \left[\delta + \frac{1}{\pi}\arccos\left(\frac{\sigma^2-1}{\sigma^2+1}\right) - 1\right]^{-2} \left[\log(2/\varepsilon) + s\log(n/s)\right].$$

Corollary 8.10. *Let* $A \in \mathbb{R}^{m\times n}$ *with rows drawn independently from the standard Gaussian distribution and* $\{\eta(x)\}$ *a collection of pairwise random vectors indexed on the sphere with each* $\eta(x) \sim \mathcal{N}(\boldsymbol{0}, \sigma^2 I_m)$. *Then for any* $0 < \varepsilon < 1$ *and* $\delta > \frac{1}{\pi}\arccos\left(\frac{1}{1+\sigma^2}\right)$, Ψ_A^η *has the* (s, δ)*-RIP with probability at least* $1 - \varepsilon$ *provided*

$$m \gtrsim \left[\delta - \frac{1}{\pi}\arccos\left(\frac{1}{1+\sigma^2}\right)\right]^{-2} \left[\log(2/\varepsilon) + s\log(n/s)\right].$$

8.2 The VC-Dimension of Sparse Hemispheres

Let X be a set and $\mathcal{C}$ be a collection of subsets of X. Denote by $\binom{X}{k}$ the set of subsets of X with k elements. For each $k \in \mathbb{N}$, define

$$m^{\mathcal{C}}(k) := \max_{B \in \binom{X}{k}} |\{B \cap C : C \in \mathcal{C}\}|.$$

Clearly $m^{\mathcal{C}}(k) \leqslant 2^k$. The *Vapnik-Chervonenkis dimension* (VC$-$dimension) of $\mathcal{C}$, denoted $VC(\mathcal{C})$, is the supremum of integers d such that $m^{\mathcal{C}}(d) = 2^d$. Alternatively, we say $\mathcal{C}$ *shatters* B if every subset of B is realized as the intersection of B with an element of $\mathcal{C}$. Then $VC(\mathcal{C})$ is the cardinality of the largest subset it shatters. For example, if $X = \mathbb{R}$ and $\mathcal{C} = \{(-\infty, t] : t \in \mathbb{R}\}$, then $VC(\mathcal{C}) = 1$; if $\mathcal{C} = \{[a, b] : a < b \in \mathbb{R}\}$, then $VC(\mathcal{C}) = 2$. VC dimension measures, in an intuitive sense, the complexity of a class of subsets.

8.2.1 VC-Dimension Background

Sauer's lemma is a fundamental result in VC theory, and we will use it several times. A proof and other details on the subject can be found in [24].

Theorem H (Sauer's Lemma). *Let $\mathcal{C}$ be a class of subsets with $\mathrm{VC}(\mathcal{C}) = d < \infty$. Then for any $k \geqslant d$,*

$$m^{\mathcal{C}}(k) \leqslant \left(\tfrac{ek}{d}\right)^d.$$

For a class of functions $\mathcal{F} \subset \{f : X \to \{0, 1\}\}$, denote by $\mathcal{C}_{\mathcal{F}}$ the set of subgraphs of functions in $\mathcal{F}$: $\mathcal{C}_{\mathcal{F}} = \{\{(x, t) : t \leqslant f(x)\} : f \in \mathcal{F}\}$. The VC dimension of $\mathcal{F}$ is defined as $\mathrm{VC}(\mathcal{C}_{\mathcal{F}})$, where this last quantity is the VC-dimension of a class of subsets of $X \times \mathbb{R}$. It is worth noting that if $\mathcal{F}$ is the set of indicators of subsets in the class $\mathcal{C}$, $\mathcal{F} = \{1_C : C \in \mathcal{C}\}$, then $\mathrm{VC}(\mathcal{F}) = \mathrm{VC}(\mathcal{C})$.

It is well known in learning theory that empirical processes in the form of (8.14) can be bounded via the VC-dimension of the indexing class. Such results are often eponymously referred to as the "VC inequality" after Vapnik and Chervonenkis, the pioneers of the theory. In Section 8.3 we use a version of the VC inequality from [69], which extends the VC inequality to a more general case, when a class satisfies *uniform entropy bounds*. For a function f and a probability $\mathbb{P}$, denote by $\mathbb{P}f$ the expectation $\int f \, d\mathbb{P}$. For a class of binary functions $\mathcal{F}$ and a probability $\mathbb{P}$, the packing number $D(\mathcal{F}, t, \mathbb{P})$ is the cardinality of the largest subset $\mathcal{F}' \subset \mathcal{F}$ such that $\mathbb{P}|f - g| > t^2$ for all $f \neq g \in \mathcal{F}'$. Finally, set

$$D(\mathcal{F}, t) := \sup_{\mathbb{P}} D(\mathcal{F}, t, \mathbb{P}),$$

where the supremum is taken over all discrete probabilities. Then [69, corollary 1] reads

Theorem I. *Suppose*

$$\int_0^\infty \sqrt{\log D(\mathcal{F}, t)}\,dt < \infty.$$

Then there exists an absolute constant $K > 0$ *such that for any* $u > 0$ *with probability at least* $1 - 2e^{-u}$ *for all* $f \in \mathcal{F}$*:*

$$\sum_{k=1}^{m} (\mathbb{P}f - f(x_k)) \leqslant K\sqrt{m}\left(\sqrt{u\mathbb{P}f} + \int_0^{\sqrt{\mathbb{P}f}} \sqrt{\log(D(\mathcal{F}, t))}\,dt\right).$$

8.2.2 Main VC Estimate

This section is dedicated to the proof of Theorem 8.1. We begin by computing the VC-dimension of all hemispheres, the case when $s = n$.

Lemma 8.11. $VC(H_s^s) = s$.

Proof. We first observe H_s^s shattering the standard basis vectors $B = \{e_1, \dots, e_s\}$, and hence $VC(H_s^s) \geqslant s$. Let $S \subset [s]$ and $B(S) = \{e_j : j \in S\}$. Define $p = (p_1, \dots, p_s)$ by setting $p_j = 1_S(j) - 1_{S^c}(j)$. Then $B(S) = B \cap H_p$.

On the other hand, let $X = \{x_1, \dots, x_{s+1}\}$ be an arbitrary $(s+1)$-subset of $\mathbb{S}^{s-1}$. Without loss of generality, assume

$$x_{s+1} = \sum_{k=1}^{s} \alpha_k x_k.$$

Set $A := \{x_k : \alpha_k < 0\} \cup \{x_{s+1}\}$; we'll see that for all $p \in \mathbb{R}^s$, $A \neq X \cap H_p$. For any p such that $\langle p, x_k \rangle > 0$ if $\alpha_k < 0$ and $\langle p, x_k \rangle \leqslant 0$ if $\alpha_k \geqslant 0$,

$$\begin{aligned}
\langle p, x_{s+1} \rangle &= \sum_{k=1}^{s} \alpha_k \langle p, x_k \rangle \\
&= \sum_{k:\alpha_k < 0} \alpha_k \langle p, x_k \rangle + \sum_{k:\alpha_k \geqslant 0} \alpha_k \langle p, x_k \rangle \\
&\leqslant 0.
\end{aligned}$$

Therefore H_s^s doesn't shatter X, so $\mathrm{VC}(H_s^s) < s+1$. $\square$

We are now ready to estimate $\mathrm{VC}(H^{n,s})$. Let $d = \mathrm{VC}(H^{n,s}) \leqslant n$ and choose a subset $X = \{x^1, \dots, x^d\}$ of $\mathbb{S}^{n-1}$ shattered by $H^{n,s}$. Fix an index set $S \in \binom{[n]}{s}$; for $x \in \mathbb{S}^{n-1}$ let $x_S = \sum_{j \in S} \langle x, e_j \rangle e_j$. For any $B \subset \mathbb{S}^{n-1}$, let $B_S = \{b_S / \|b_S\| : b \in B \text{ and } b_S \neq 0\}$. Notice that $|X_S| \leqslant d$, so by Lemmas 8.11 and Theorem H,

$$\left|\{X_S \cap H_p : p \in \mathbb{S}_S^{n-1}\}\right| \leqslant \left(\tfrac{ed}{s}\right)^s.$$

The natural map $\{X \cap H_p : p \in \mathbb{S}_S^{n-1}\} \to \{X_S \cap H_p : p \in \mathbb{S}_S^{n-1}\}$ via $A \mapsto A_S$ is well-defined and surjective since $\mathrm{sgn}(\langle x, p \rangle) = \mathrm{sgn}(\langle x_S, p \rangle)$ for all $x \in X$ and $p \in \mathbb{S}_S^{n-1}$. This map is also injective. Suppose $A = X \cap H_p$ and $B = X \cap H_{p'}$ are distinct, for instance $a \in A \setminus B$ (hence $a_S \neq 0$). If $a_S/\|a_S\| \in B_S$, then there is $b \in B$ such that $a_S/\|a_S\| = b_S/\|b_S\|$. But then $\mathrm{sgn}(\langle a, p' \rangle) = \mathrm{sgn}(\langle b, p' \rangle)$, a contradiction.

It follows that

$$\left|\{X \cap H_p : p \in \mathbb{S}_S^{n-1}\}\right| \leqslant \left(\tfrac{ed}{s}\right)^s,$$

and by the union bound, $2^d \leqslant \binom{n}{s}\left(\frac{ed}{s}\right)^s$. After applying a familiar version of Stirlings approximation, $\binom{n}{s} \leqslant \left(\frac{en}{s}\right)^s$, and some algebraic manipulation, we arrive at the inequality:

$$-\log(2)\tfrac{d}{s} e^{-\log(2)\frac{d}{s}} \leqslant -\tfrac{\log(2)s}{e^2 n}. \tag{8.12}$$

To simplify further, we use the lower branch of the *Lambert W function*, which is defined on $(\frac{-1}{e}, 0)$ by the relation $W_{-1}(x) e^{W_{-1}(x)} = x$. That is, W_{-1} is the inverse of the map $x \mapsto xe^x$ restricted to $(-\infty, -1)$. We use the following lower bound of W_{-1} to simplify (8.12).

Lemma 8.13. *For all $-1/e < x < 0$, $W_{-1}(x) \geqslant \log(x^2)$.*

Proof. Notice that W_{-1} is decreasing, as is its inverse $W_{-1}^{-1}(x) = xe^x$. Applying W_{-1}^{-1} to each side of the equation in the statement and dividing by x, we find the equivalent: $x\log(x^2) \leqslant 1$ for all $\frac{-1}{e} < x < 0$. This holds since $x \mapsto x\log(x^2)$ is decreasing on $(-1/e, 0)$ and $(\frac{-1}{e})\log(\frac{1}{e^2}) = \frac{2}{e} < 1$. □

Applying the decreasing W_{-1} to both sides of (8.12) and using Lemma 8.13 gives:

$$d \leqslant \tfrac{2}{\log(2)} s \log\left(\tfrac{ne^2}{s\log(2)}\right).$$

8.3 The RIP of one-bit Embeddings

This section proves Theorem 8.3. Let $A \in \mathbb{R}^{m\times n}$ with rows $\{g_k\}_{k=1}^m$ drawn independently from the standard Gaussian distribution $\mathcal{N}(\mathbf{0}, I_n)$. The Hamming distance between the images of two signals x and y under the one-bit embedding Φ_A is

$$d_{\mathcal{H}^m}(\Phi_A x, \Phi_A y) = \frac{1}{m}\sum_{k=1}^{m} \frac{1 - \operatorname{sgn}\langle x, g_k\rangle \operatorname{sgn}\langle y, g_k\rangle}{2}.$$

For $x, y \in \mathbb{S}^{n-1}$ we call $W_{x,y} := H_x \triangle H_y$ (the symmetric difference of the two hemispheres) the *wedge* associated to x and y. Notice $\operatorname{sgn}\langle x, g_k\rangle \neq \operatorname{sgn}\langle y, g_k\rangle$ if and only if g_k is in the wedge $W_{x,y}$. The Hamming distance above can be reformulated as

$$d_{\mathcal{H}^m}(\Phi_A x, \Phi_A y) = \frac{1}{m}\sum_{k=1}^{m} 1_{W_{x,y}}(g_k).$$

The empirical processes framework suggests the sphere should be endowed with the distance $(x, y) \mapsto \mathbb{P}(W_{x,y})$. Fix $x, y \in \mathbb{S}^{n-1}$, let $g \sim \mathcal{N}(\mathbf{0}, I_n)$, and let $Z = (\langle x, g\rangle, \langle y, g\rangle)^\top$; then $Z \sim \mathcal{N}(\mathbf{0}, \Sigma)$ with

$$\Sigma = \begin{bmatrix} 1 & \langle x, y\rangle \\ \langle x, y\rangle & 1 \end{bmatrix}.$$

It is a basic computation to find

$$\begin{aligned}\mathbb{P}(W_{x,y}) &= \tfrac{1}{\pi\sqrt{1-\langle x,y\rangle^2}} \int_0^\infty \int_0^\infty \operatorname{Exp}\left(\tfrac{2uv\langle x,y\rangle - u^2 - v^2}{2-2\langle x,y\rangle^2}\right) du\, dv \\ &= \tfrac{1}{\pi} \arccos\left(\langle x,y\rangle\right).\end{aligned}$$

This last quantity is the geodesic distance on the sphere that we denote by $d(x,y)$. This brings our attention to the following object:

$$\sup_{x,y \in \mathbb{S}_s^{n-1}} \left| \frac{1}{m} \sum_{k=1}^{m} 1_{W_{x,y}}(g_k) - d(x,y) \right|. \tag{8.14}$$

The above formulation is paraphrased from [4]; this is the point at which our argument deviates. To utilize the VC theory for hemispheres developed in the previous section, we bound the VC-dimension of the class of "sparse wedges" $\mathcal{W}^{n,s} := \{W_{x,y} : x, y \in \mathbb{S}_s^{n-1}\}$.

Lemma 8.15. *Let $\mathcal{C}$ be a class of subsets of X with $\mathrm{VC}(\mathcal{C}) = d < \infty$. Let $\mathcal{C}\triangle\mathcal{C} = \{C \triangle C' : C, C' \in \mathcal{C}\}$. Then $\mathrm{VC}(\mathcal{C}\triangle\mathcal{C}) \leqslant 10d$.*

Proof. Let $B \subset X$ of size $m := |B|$ to be prescribed later. For a fixed pair $C, C' \in \mathcal{C}$, notice that

$$B \cap (C \triangle C') = [(B \cap C) \setminus (B \cap C')] \cup [(B \cap C') \setminus (B \cap C)].$$

That is, $B \cap (C \triangle C')$ is determined by $B \cap C$ and $B \cap C'$. By Theorem H, there are no more than $\left(\frac{em}{d}\right)^{2d}$ such pairs. Taking $m \geqslant 10d$ yields $\left(\frac{em}{d}\right)^{2d} < 2^m$. □

Along with Theorem 8.1, this lemma implies $\mathrm{VC}(\mathcal{W}^{n,s}) \lesssim s \log(n/s)$. We use this VC-dimension estimate to bound the *packing numbers* of the sparse wedges, $D(\mathcal{W}^{n,s}, \varepsilon, \mathbb{P})$, which is the largest d so that there exists $w_1, \dots, w_d \in \mathcal{W}^{n,s}$ with $\mathbb{P}(w_i \triangle w_j) > \varepsilon^2$ for $i \neq j$. General results bounding packing numbers via VC-dimension are well-known and the argument is standard; we include a proof in the current context for completeness.

Proposition 8.16. *For* $0 < \varepsilon < 1$,

$$D(\mathcal{W}^{n,s}, \varepsilon, \mathbb{P}) \lesssim \left(\tfrac{1}{\varepsilon^2}\right)^{VC(\mathcal{W}^{n,s})+1}.$$

Proof. Fix $0 < \varepsilon < 1$. Let $d = D(\mathcal{W}^{n,s}, \varepsilon, \mathbb{P})$ and let $w_1 \dots, w_d$ such that $\mathbb{P}(w_i \triangle w_j) > \varepsilon^2$ for all $i \neq j$. Let $\{X_k\}_{k=1}^n$ be independent and identically distributed on the sphere with law $\mathbb{P}$, where n will be determined later. Notice that $w_i \cap \{X_k\} \neq w_j \cap \{X_k\}$ if and only if $(w_i \triangle w_j) \cap \{X_k\}$ is nonempty. Thus the probability that there is $i \neq j$ such that $w_i \cap \{X_k\} = w_j \cap \{X_k\}$ is no more than

$$\begin{aligned}
\binom{d}{2} \max_{1 \leqslant i \neq j \leqslant d} \mathbb{P}\left(w_i \cap \{X_k\} = w_j \cap \{X_k\}\right) &= \binom{d}{2} \max_{1 \leqslant i \neq j \leqslant d} \left(1 - \mathbb{P}(w_i \triangle w_j)\right)^n \\
&< \binom{d}{2}(1-\varepsilon^2)^n \\
&< d^2 e^{-n\varepsilon^2} \\
&= e^{2\log(d) - n\varepsilon^2}.
\end{aligned}$$

Now we take $n = \left\lceil \frac{2\log(d)+1}{\varepsilon^2} \right\rceil$ so the above probability is less than one, hence there is a deterministic $X = \{x_k\}_{k=1}^n$ so that the intersections $\{w_j \cap X\}_{j=1}^d$ are distinct. Let $\nu = VC(\mathcal{W}^{n,s})$. Employing Theorem H, there is $K_\nu > 0$ such that

$$d \leqslant K_\nu \left(\frac{2\log(d)+2}{\varepsilon^2}\right)^\nu.$$

Choose d_0 large enough so that for $d > d_0$, $(2\log(d)+2)^{\nu+1} < d^{1/\nu}$. This yields

$$d \leqslant \max\{d_0, K_\nu^{\frac{\nu+1}{\nu}}\} \left(\tfrac{1}{\varepsilon^2}\right)^{(\nu+1)}.$$

□

Notice that the bound in Proposition 8.16 holds uniformly over all probabilities on the sphere. This fact allows us to use a version of the entropy integral in the final stage of our argument. Recall Theorem I. Adapted to our current setting, we have the following corollary:

Corollary 8.17. *There exists an absolute constant* $K > 0$ *such that for any* $u > 0$ *with probability at least* $1 - 2e^{-u}$ *for all* $W_{x,y} \in \mathcal{W}^{n,s}$*:*

$$\sum_{k=1}^{m} \left(d(x,y) - 1_{W_{x,y}}(g_k) \right) \leqslant K\sqrt{m} \left(\sqrt{u d(x,y)} + \int_0^{\sqrt{d(x,y)}} \sqrt{\log(D(\mathcal{W}^{n,s}, t))} dt \right).$$

We adjust this result in two ways to produce the main results of this section. First, increase the right side of the inequality by replacing all distances with one. Now that the bound is uniform over pairs of signals in $\mathbb{S}_s^{n-1}$, we observe

$$\sup_{x,y \in \mathbb{S}_s^{n-1}} \sum_{k=1}^{m} \left(d(x,y) - 1_{W_{x,y}}(g_k) \right) = \sup_{x,y \in \mathbb{S}_s^{n-1}} \left| \sum_{k=1}^{m} 1_{W_{x,y}}(g_k) - d(x,y) \right|.$$

This is because $1_{W_{-x,y}} = 1 - 1_{W_{x,y}}$ (a.s.), and $d(-x,y) = 1 - d(x,y)$. Thus we have:

Corollary 8.18. *There exists an absolute constant* $K > 0$ *such that for any* $u > 0$ *with probability at least* $1 - 2e^{-u}$,

$$\sup_{x,y \in \mathbb{S}_s^{n-1}} \frac{1}{m} \left| \sum_{k=1}^{m} 1_{W_{x,y}}(g_k) - d(x,y) \right| \leqslant \frac{K}{\sqrt{m}} \left(\sqrt{u} + \int_0^1 \sqrt{\log(D(\mathcal{W}^{n,s}, t))} dt \right).$$

After applying the uniform entropy bounds of Proposition 8.16 in the above corollary and setting $u = \log(2/\varepsilon)$, Theorem 8.3 is immediate.

8.4 The RIP of Noisy one-bit Embeddings

8.4.1 Systematically Noisy RIP with the Distorted Metric

This section proves Theorem 8.7. We again consider $A \in \mathbb{R}^{m\times n}$ with rows $\{g_k\}_{k=1}^m$ drawn independently from the standard Gaussian distribution $\mathcal{N}(\mathbf{0}, I_n)$. We are now interested in the case of systematic additive white noise prior to quantization; let $\eta \sim \mathcal{N}(\mathbf{0}, \sigma^2 I_m)$. Then the Hamming distance between the images of two signals under the one-bit embedding Φ_A^η is

$$d_{\mathcal{H}^m}(\Phi_A^\eta x, \Phi_A^\eta y) = \frac{1}{m}\sum_{k=1}^{m} \frac{1 - \operatorname{sgn}(\langle x, g_k\rangle + \eta_k)\operatorname{sgn}(\langle y, g_k\rangle + \eta_k)}{2}. \qquad (8.19)$$

Fix $x, y \in \mathbb{S}^{n-1}$. Let $g \sim \mathcal{N}(\mathbf{0}, I_n)$ and $\mu \sim \mathcal{N}(0, \sigma^2)$ be independent. Then $\binom{\langle x,g\rangle+\mu}{\langle y,g\rangle+\mu}$ is a Gaussian vector with covariance matrix

$$\begin{bmatrix} 1+\sigma^2 & \langle x,y\rangle+\sigma^2 \\ \langle x,y\rangle+\sigma^2 & 1+\sigma^2 \end{bmatrix}.$$

A computation similar to 8.3 yields

$$\mathbb{P}\big(\operatorname{sgn}(\langle x,g\rangle+\mu)\operatorname{sgn}(\langle y,g\rangle+\mu) = -1\big) = \tfrac{1}{\pi}\arccos\left(\tfrac{\langle x,y\rangle+\sigma^2}{1+\sigma^2}\right).$$

This last quantity is $d^\sigma(x, y)$, defined in (8.4). We'll see soon that d^σ is in fact a metric; this is the distance with which $\mathbb{S}_s^{n-1}$ is naturally endowed in the presence of systematic additive white noise. The object in the $(s, \delta)^\sigma$-RIP that we aim to bound is

$$\sup_{x,y\in\mathbb{S}_s^{n-1}} \left| d_{\mathcal{H}^m}(\Phi_A^\eta x, \Phi_A^\eta y) - d^\sigma(x,y)\right|.$$

Appealing to the methods in Section 8.3, we rewrite the additive noise as an inner product by increasing the Gaussian measurements' dimension by one and

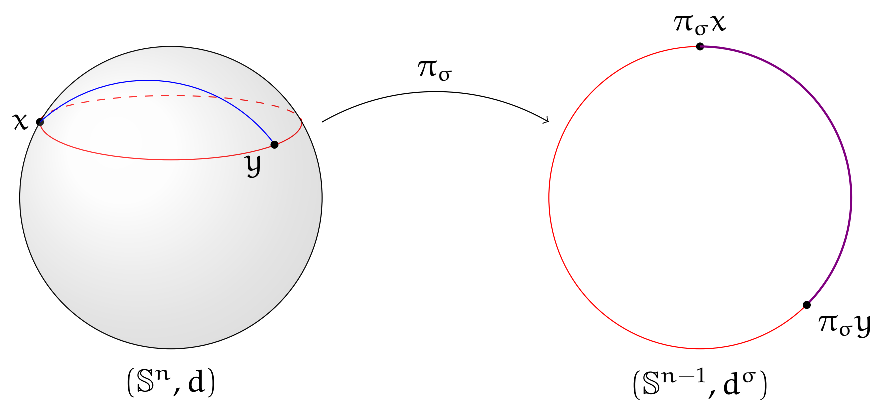

Figure 8.1: If $\pi_\sigma : \{p \in \mathbb{S}^n : \langle p, e_{n+1}\rangle = \sigma\} \to \mathbb{S}^{n-1}$ is the normalization of the projection onto the first n coordinates, then $d^\sigma(\pi_\sigma x, \pi_\sigma y) = d(x, y)$.

lifting the sphere to one higher dimension by padding with σ. Introduce the following notation:

$$x_\sigma = \tfrac{1}{\sqrt{1+\sigma^2}}(x^1, \dots, x^n, \sigma) \in \mathbb{S}^n_{s+1}.$$

Let $h = (g^1, \dots, g^n, \frac{1}{\sigma}\eta)$ and notice $\langle x_\sigma, h\rangle = \frac{1}{\sqrt{1+\sigma^2}}(\langle x, g\rangle + \eta)$ and $h \sim \mathcal{N}(\mathbf{0}, I_{n+1})$. Denote by $W^\sigma_{x,y}$ the wedge in $\mathbb{S}^n$ relative to x_σ and y_σ, i.e.,

$$W^\sigma_{x,y} := H_{x_\sigma} \triangle H_{y_\sigma}.$$

Then $\mathrm{sgn}(\langle x, g_k\rangle + \eta_k) \neq \mathrm{sgn}(\langle y, g_k\rangle + \eta_k)$ if and only if $h_k := (g^1_k, \dots, g^n_k, \frac{1}{\sigma}\eta_k) \in W^\sigma_{x,y}$. The Hamming distance in (8.20) can be reformulated as

$$d_{\mathcal{H}^m}(\Phi^\eta_A x, \Phi^\eta_A y) = \frac{1}{m}\sum_{k=1}^{m} 1_{W^\sigma_{x,y}}(h_k).$$

Furthermore, notice that

$$\langle x_\sigma, y_\sigma\rangle = \tfrac{\langle x, y\rangle + \sigma^2}{1+\sigma^2},$$

hence $d^\sigma(x, y) = d\,(x_\sigma, y_\sigma)\,$, where we abuse notation to allow $d\,(\cdot, \cdot)$ to denote the normalized geodesic distance on $\mathbb{S}^n$; see Figure 8.1 for an illustration. It is now apparent that d^σ is indeed a metric on $\mathbb{S}^{n-1}$.

This brings our attention to the following object:

$$\sup_{x,y\in\mathbb{S}_s^{n-1}} \left| \frac{1}{m}\sum_{k=1}^{m} 1_{W^\sigma_{x,y}}(h_k) - d^\sigma(x,y) \right|,$$

where the row vectors h_k are independently drawn from $\mathcal{N}(\mathbf{0}, I_{n+1})$. At this point, the argument of Section 8.3 applies so long as we can estimate the VC-dimension of

$$\mathcal{W}_\sigma^{n,s} := \{W_{x,y} \in \mathcal{W}^{n+1,s+1} : x^{n+1} = \sigma = y^{n+1}\}.$$

Since $\mathcal{W}_\sigma^{n,s} \subset \mathcal{W}^{n+1,s+1}$, it is clear that

$$\mathrm{VC}(\mathcal{W}_\sigma^{n,s}) \leqslant \mathrm{VC}(\mathcal{W}^{n+1,s+1}) \lesssim (s+1)\log\left(\tfrac{n+1}{s+1}\right) \lesssim s\log(n/s).$$

8.4.2 Independently Noisy RIP with the Distorted Metric

We extend the results of the previous section to prove Theorem 8.8. We still consider $A \in \mathbb{R}^{m\times n}$ with rows $\{g_k\}_{k=1}^m$ drawn independently from the standard Gaussian distribution $\mathcal{N}(\mathbf{0}, I_n)$. However, we are now interested in the case of *independent* additive white noise prior to quantization; let $\eta(x) \sim \mathcal{N}(\mathbf{0}, \sigma^2 I_m)$ for each $x \in \mathbb{S}^{n-1}$ with $\eta(x), \eta(y)$ independent when x and y are distinct. Let $g \sim \mathcal{N}(\mathbf{0}, I_n)$ and $\mu_x, \mu_y \sim \mathcal{N}(0, \sigma^2)$ be mutually independent. Then the Hamming distance between the images of two signals under this one-bit embedding Ψ_A^η is

$$d_{\mathcal{H}^m}(\Psi_A^\eta x, \Psi_A^\eta y) = \frac{1}{m}\sum_{k=1}^{m} \frac{1-\operatorname{sgn}\left[(\langle x, g_k\rangle + \eta_k(x))(\langle y, g_k\rangle + \eta_k(y))\right]}{2}. \tag{8.20}$$

Fix $x, y \in \mathbb{S}^{n-1}$. Then $\binom{\langle x,g\rangle+\mu_x}{\langle y,g\rangle+\mu_y}$ is a Gaussian vector with covariance matrix

$$\begin{bmatrix} 1+\sigma^2 & \langle x,y\rangle \\ \langle x,y\rangle & 1+\sigma^2 \end{bmatrix},$$

and it follows that

$$\mathbb{P}\big(\mathrm{sgn}(\langle x,g\rangle+\mu)\mathrm{sgn}(\langle y,g\rangle+\mu)=-1\big)=\tfrac{1}{\pi}\arccos\left(\tfrac{\langle x,y\rangle}{1+\sigma^2}\right).$$

The last quantity is $d_\sigma(x,y)$, defined in (8.5). d_σ is also a metric; this is the distance with which $\mathbb{S}_s^{n-1}$ is naturally endowed in the presence of independent additive white noise. To proceed, we fix x, y on the sphere and lift them to $\mathbb{S}_{s+2}^{n+1}$:

$$x_{0,\sigma}=\tfrac{1}{\sqrt{1+\sigma^2}}(x^1,\dots,x^n,0,\sigma)$$
$$y_{\sigma,0}=\tfrac{1}{\sqrt{1+\sigma^2}}(x^1,\dots,x^n,\sigma,0).$$

After increasing the dimension of the Gaussian measurements by two, the remainder of the proof is completely analogous to that of Theorem 8.7.

8.4.3 Noisy RIPs with geodesic metric on the sphere

The deviation of d^σ from the geodesic distance is exaggerated at antipodes. That is, for any x and y on the sphere, $|d(x,y)-d^\sigma(x,y)|\leqslant d(x,-x)-d^\sigma(x,-x)$. In what is to come, all suprema are over $x, y \in \mathbb{S}_s^{n-1}$. If one prefers a bound of the form

$$\sup|d_{\mathcal{H}^m}(\Phi_A^\eta x,\Phi_A^\eta y)-d(x,y)|\leqslant\delta,$$

it is enough for

$$\begin{aligned}\sup|d_{\mathcal{H}^m}(\Phi_A^\eta x, \Phi_A^\eta y) - d(x,y)| &\leqslant \sup|d_{\mathcal{H}^m}(\Phi_A^\eta x, \Phi_A^\eta y) - d^\sigma(x,y)| + |d^\sigma(x,y) - d(x,y)| \\ &\leqslant \sup|d_{\mathcal{H}^m}(\Phi_A^\eta x, \Phi_A^\eta y) - d^\sigma(x,y)| + 1 - \tfrac{1}{\pi}\arccos\left(\tfrac{\sigma^2-1}{\sigma^2+1}\right) \\ &\leqslant \delta.\end{aligned}$$

Corollary 8.9 follows easily from Theorem 8.7 and this observation. Similarly, we have

$$\begin{aligned}\sup|d_{\mathcal{H}^m}(\Psi_A^\eta x, \Psi_A^\eta y) - d(x,y)| &\leqslant \sup|d_{\mathcal{H}^m}(\Psi_A^\eta x, \Psi_A^\eta y) - d_\sigma(x,y)| + |d_\sigma(x,y) - d(x,y)| \\ &\leqslant \sup|d_{\mathcal{H}^m}(\Psi_A^\eta x, \Psi_A^\eta y) - d_\sigma(x,y)| + \tfrac{1}{\pi}\arccos\left(\tfrac{1}{1+\sigma^2}\right),\end{aligned}$$

which yields Corollary 8.10

REFERENCES

[1] L. Anitori, A. Maleki, M. Otten, R. Baraniuk, and P. Hoogeboom, "Design and analysis of compressed sensing radar detectors," *IEEE Trans. Signal Process.*, vol. 61, no. 4, pp. 813–827, 2013.

[2] C. Benea, F. Bernicot, and T. Luque, "Sparse bilinear forms for Bochner-Riesz multipliers and applications," eprint: `http://arxiv.org/abs/1605.06401`.

[3] F. Bernicot, D. Frey, and S. Petermichl, "Sharp weighted norm estimates beyond Calderón–Zygmund theory," *Anal. PDE*, vol. 9, no. 5, pp. 1079–1113, 2016.

[4] D. Bilyk and M. T. Lacey, "Random tessellations, restricted isometric embeddings, and one bit sensing," *ArXiv e-prints*, 2015. eprint: `https://arxiv.org/abs/1512.06697`.

[5] S. Bloom, "A commutator theorem and weighted bmo," *Trans. Amer. Math. Soc.*, vol. 292, no. 1, pp. 103–122, 1985.

[6] J. Bobin, J. Starck, and R. Ottensamer, "Compressed sensing in astronomy," *IEEE Sel. Top. Signal Proc.*, vol. 2, p. 718, 2008.

[7] P. Boufounos and R. Baraniuk, "1-bit compressive sensing," in *Proceedings of Conference on Information Science and Systems (CISS), Princeton, NJ*, March 2008.

[8] J. Bourgain, "On the maximal ergodic theorem for certain subsets of the integers," *Israel J. Math.*, vol. 61, no. 1, pp. 39–72, 1988.

[9] Z. Buczolich and R. D. Mauldin, "Divergent square averages," *Ann. of Math. (2)*, vol. 171, no. 3, pp. 1479–1530, 2010.

[10] E. Candés, "The restricted isometry property and its implications for compressed sensing," *C. R. Math. Acad. Sci. Paris*, vol. Ser. I 346, 2008.

[11] L. Chaffee, "Characterizations of BMO through commutators of bilnear singular integral operators," 2014.

[12] S. Chanillo, "A note on commutators," *Indiana Univ. Math. J.*, vol. 31, no. 1, pp. 7–16, 1982.

[13] S. Chanillo and M. Christ, “Weak $(1, 1)$ bounds for oscillatory singular integrals,” *Duke Math. J.*, vol. 55, no. 1, pp. 141–155, 1987.

[14] Coifman and Fefferman, “Weighted norm inequalities for maximal functions and singular integrals,” 1974.

[15] R. R. Coifman and C. Fefferman, “Weighted norm inequalities for maximal functions and singular integrals,” *Studia Math.*, vol. 51, 241–250, 1974.

[16] J. M. Conde-Alonso and G. Rey, “A pointwise estimate for positive dyadic shifts and some applications,” *Math. Ann.*, vol. 365, no. 3-4, pp. 1111–1135, 2016.

[17] D. Cruz-Uribe, J. M. Martell, and C. Pérez, “Sharp two-weight inequalities for singular integrals, with applications to the Hilbert transform and the Sarason conjecture,” *Adv. Math.*, vol. 216, no. 2, pp. 647–676, 2007.

[18] D. Cruz-Uribe, “Two weight norm inequalities for fractional integral operators and commutators,” 2015. eprint: `http://arxiv.org/abs/1412.4157`.

[19] D. Cruz-Uribe and K. Moen, “A fractional Muckenhoupt-Wheeden theorem and its consequences,” *Integral Equations Operator Theory*, vol. 76, no. 3, pp. 421–446, 2013.

[20] ——, “One and two weight norm inequalities for Riesz potentials,” *Illinois J. Math.*, vol. 57, no. 1, pp. 295–323, 2013.

[21] D. Cruz-Uribe, A. Reznikov, and A. Volberg, “Logarithmic bump conditions and the two-weight boundedness of Calderón-Zygmund operators,” *Adv. Math.*, vol. 255, pp. 706–729, 2014.

[22] D. V. Cruz-Uribe, J. M. Martell, and C. Pérez, *Weights, extrapolation and the theory of rubio de francia*, ser. Operator Theory: Advances and Applications. Birkhäuser/Springer Basel AG, Basel, 2011, vol. 215, pp. xiv+280.

[23] A. Culiuc, F. Di Plinio, and Y. Ou, “Domination of multilinear singular integrals by positive sparse forms,” eprint: `http://arxiv.org/abs/1603.05317`.

[24] A. DasGupta, *Probability for statistics and machine learning: Fundamentals and advanced topics*, 1st. Springer Publishing Company, Incorporated, 2011, ISBN: 1441996338, 9781441996336.

[25] F. C. de França Silva and P. Zorin-Kranich, "Sparse domination of noncentered variational operators," eprint: http://arxiv.org/abs/1604.05506.

[26] S. Ding, "Two-weight caccioppoli inequalities for solutions of nonhomogeneous a-harmonic equations on riemannian manifolds," *Proc. Amer. Math. Soc.*, vol. 132, no. 8, pp. 2367–2375, 2004.

[27] Y. Ding and H. Liu, "Uniform weighted estimates for oscillatory singular integrals," *Forum Math.*, vol. 24, no. 2, pp. 223–238, 2012.

[28] ——, "Weighted L^p boundedness of Carleson type maximal operators," *Proc. Amer. Math. Soc.*, vol. 140, no. 8, pp. 2739–2751, 2012.

[29] P. L. Duren, "Extension of a theorem of Carleson," *Bull. Amer. Math. Soc.*, vol. 75, pp. 143–146, 1969.

[30] M. Folch-Gabayet and J. Wright, "Weak-type $(1,1)$ bounds for oscillatory singular integrals with rational phases," *Studia Math.*, vol. 210, no. 1, pp. 57–76, 2012.

[31] L. Grafakos, J. M. Martell, and F. Soria, "Weighted norm inequalities for maximally modulated singular integral operators," *Math. Ann.*, vol. 331, no. 2, pp. 359–394, 2005.

[32] I. Holmes, M. T. Lacey, and B. D. Wick, "Commutators in the two-weight setting," *ArXiv preprint arXiv:1506.05747*, 2015.

[33] I. Holmes, M. T. Lacey, and B. D. Wick, "Bloom's inequality: Commutators in a two-weight setting," 2015. eprint: http://arxiv.org/abs/1505.07947.

[34] R. Hunt, B. Muckenhoupt, and R. Wheeden, "Weighted norm inequalities for the conjugate function and Hilbert transform," *Trans. Amer. Math. Soc.*, vol. 176, pp. 227–251, 1973.

[35] T. P. Hytönen, "On Petermichl's dyadic shift and the Hilbert transform," *C. R. Math. Acad. Sci. Paris*, vol. 346, no. 21-22, pp. 1133–1136, 2008.

[36] ——, "The A_2 theorem: Remarks and complements," in *Harmonic analysis and partial differential equations*, ser. Contemp. Math. Vol. 612, Amer. Math. Soc., Providence, RI, 2014, pp. 91–106.

[37] T. P. Hytönen and M. T. Lacey, "The A_p-A_∞ inequality for general Calderón-Zygmund operators," *Indiana Univ. Math. J.*, vol. 61, no. 6, pp. 2041–2092, 2012.

[38] T. P. Hytönen, M. T. Lacey, H. Martikainen, T. Orponen, M. C. Reguera, E. T. Sawyer, and I. Uriarte-Tuero, "Weak and strong type estimates for maximal truncations of Calderón-Zygmund operators on A_p weighted spaces," *J. Anal. Math.*, vol. 118, no. 1, pp. 177–220, 2012.

[39] T. P. Hytönen and C. Pérez, "Sharp weighted bounds involving A_∞," *Anal. PDE*, vol. 6, no. 4, pp. 777–818, 2013.

[40] T. P. Hytönen, C. Pérez, S. Treil, and A. Volberg, "Sharp weighted estimates for dyadic shifts and the A_2 conjecture," *J. Reine Angew. Math.*, vol. 687, pp. 43–86, 2014.

[41] T. P. Hytönen, L. Roncal, and O. Tapiola, "Quantitative weighted estimates for rough homogeneous singular integrals," eprint: `http://arxiv.org/abs/1510.05789`.

[42] L. Jacques and K. Degraux, "Quantized iterative hard thresholding: Bridging 1-bit and high-resolution quantized compressed sensing," 2013.

[43] L. Jacques, J. Laska, P. Boufounos, and R. Baraniuk, "Robust 1-bit compressive sensing via binary stable embeddings of sparse vectors," *IEEE Trans. Inform. Theory*, vol. 59, no. 4, pp. 2082–2102, 2013.

[44] R. Johnson and C. J. Neugebauer, "Change of variable results for A_p- and reverse hölder RH_r-classes," *Trans. Amer. Math. Soc.*, vol. 328, no. 2, pp. 639–666, 1991.

[45] B. Krause and M. T. Lacey, "A weak type inequality for maximal monomial oscillatory Hilbert transforms," eprint: `1609.01564`.

[46] M. T. Lacey, "An A_p-A_∞ inequality for the Hilbert transform," *Houston J. Math.*, vol. 38, no. 3, pp. 799–814, 2012.

[47] ——, "An elementary proof of the A_2 bound," *Israel J. Math., to appear*, 2015. eprint: `http://arxiv.org/abs/1501.05818`.

[48] ——, "Haar shifts, commutators, and hankel operators," *Rev. Un. Mat. Argentina*, vol. 50, no. 2, pp. 1–13, 2009.

[49] ——, "On the separated bumps conjecture for Calderón-Zygmund operators," *Hokkaido Math. J.*, vol. 45, no. 2, pp. 223–242, 2016.

[50] M. T. Lacey, S. Petermichl, and M. C. Reguera, "Sharp A_2 inequality for haar shift operators," *Math. Ann.*, vol. 348, no. 1, pp. 127–141, 2010.

[51] M. T. Lacey, E. T. Sawyer, and I. Uriarte-Tuero, "Two weight inequalities for discrete positive operators," *Submitted*, 2009. eprint: `http://www.arxiv.org/abs/0911.3437`.

[52] M. T. Lacey and S. Spencer, "On entropy bumps for Calderón-Zygmund operators," *Concr. Oper.*, vol. 2, pp. 47–52, 2015.

[53] P. LaVictoire, "An L^1 ergodic theorem for sparse random subsequences," *Math. Res. Lett.*, vol. 16, no. 5, pp. 849–859, 2009.

[54] A. K. Lerner and F. Nazarov, "Intuitive dyadic calculus: The basics," eprint: `http://arxiv.org/abs/1508.05639`.

[55] A. K. Lerner, S. Ombrosi, and I. P. Rivera-Ríos, "On pointwise and weighted estimates for commutators of Calderón-Zygmund operators," eprint: `http://arxiv.org/abs/1604.01334`.

[56] A. K. Lerner, "A pointwise estimate for the local sharp maximal function with applications to singular integrals," *Bull. Lond. Math. Soc.*, vol. 42, no. 5, pp. 843–856, 2010.

[57] ——, "A simple proof of the A_2 conjecture," *Int. Math. Res. Not. IMRN*, no. 14, pp. 3159–3170, 2013.

[58] ——, "Mixed A_p-A_r inequalities for classical singular integrals and littlewood-paley operators," *J. Geom. Anal.*, vol. 23, no. 3, pp. 1343–1354, 2013.

[59] ——, "On an estimate of Calderón-Zygmund operators by dyadic positive operators," *J. Anal. Math.*, vol. 121, pp. 141–161, 2013.

[60] ——, "On pointwise estimates involving sparse operators," *New York J. Math.*, vol. 22, pp. 341–349, 2016.

[61] A. K. Lerner and K. Moen, "Mixed A_p-A_∞ estimates with one supremum," *Studia Math.*, vol. 219, no. 3, pp. 247–267, 2013.

[62] M. Lustig, D. D. L., and J. M. Pauly, "Sparse MRI: The application of compressed sensing for rapid MR imaging," *Magn Reson Med*, vol. 58, pp. 1182–1195, 2007.

[63] M. Mirek, "Weak type $(1,1)$ inequalities for discrete rough maximal functions," *J. Anal. Math.*, vol. 127, pp. 247–281, 2015.

[64] B. Muckenhoupt and R. L. Wheeden, "Weighted bounded mean oscillation and the Hilbert transform," *Studia Math.*, vol. 54, no. 3, pp. 221–237, 1975–76.

[65] ——, "Weighted norm inequalities for fractional integrals," *Trans. Amer. Math. Soc.*, vol. 192, pp. 261–274, 1974.

[66] B. Muckenhoupt, "Weighted norm inequalities for the Hardy maximal function," *Trans. Amer. Math. Soc.*, vol. 165, pp. 207–226, 1972.

[67] F. Nazarov, A. Reznikov, S. Treil, and A. Volberg, "A Bellman function proof of the L^2 bump conjecture," *J. Anal. Math.*, vol. 121, pp. 255–277, 2013.

[68] C. J. Neugebauer, "Inserting A_p-weights," *Proc. Amer. Math. Soc.*, vol. 87, no. 4, pp. 644–648, 1983.

[69] D. Panchenko, "Some extensions of an inequality of Vapnik and Chervonenkis," *Electron. Comm. in Probab.*, vol. 7, 6:55–6:65, 2002.

[70] C. Pérez, "On sufficient conditions for the boundedness of the Hardy-Littlewood maximal operator between weighted L^p-spaces with different weights," *Proc. London Math. Soc. (3)*, vol. 71, no. 1, pp. 135–157, 1995.

[71] ——, "Two weighted inequalities for potential and fractional type maximal operators," *Indiana Univ. Math. J.*, vol. 43, no. 2, pp. 663–683, 1994.

[72] ——, "Weighted norm inequalities for singular integral operators," *J. London Math. Soc. (2)*, vol. 49, no. 2, pp. 296–308, 1994.

[73] S. Petermichl, "The sharp bound for the Hilbert transform on weighted lebesgue spaces in terms of the classical A_p characteristic," *Amer. J. Math.*, vol. 129, no. 5, pp. 1355–1375, 2007.

[74] S. Petermichl, S. Treil, and A. Volberg, "Why the Riesz transforms are averages of the dyadic shifts?" *Publ. Mat.*, 2002.

[75] S. Petermichl, "Dyadic shifts and a logarithmic estimate for hankel operators with matrix symbol," *C. R. Acad. Sci. Paris Sér. I Math.*, vol. 330, no. 6, pp. 455–460, 2000.

[76] Y. Plan and R. Vershynin, "Dimension reduction by random hyperplane tessellations," *Electron. Comm. in Probab.*, vol. 7, pp. 55–65, 2002.

[77] R. Rahm and S. Spencer, "Some entropy bump conditions for fractional maximal and integral operators," 2015. eprint: `http://arxiv.org/abs/1504.05906v2`.

[78] F. Ricci and E. M. Stein, "Oscillatory singular integrals and harmonic analysis on nilpotent groups," *Proc. Nat. Acad. Sci. U.S.A.*, vol. 83, no. 1, pp. 1–3, 1986.

[79] F. Ricci and E. M. Stein, "Harmonic analysis on nilpotent groups and singular integrals. i. oscillatory integrals," *J. Funct. Anal.*, vol. 73, no. 1, pp. 179–194, 1987.

[80] R. Rochberg, "NWO sequences, weighted potential operators, and Schrödinger eigenvalues," *Duke Math. J.*, vol. 72, no. 1, pp. 187–215, 1993.

[81] J. M. Rosenblatt and M. Wierdl, "Pointwise ergodic theorems via harmonic analysis," *In Ergodic theory and its connections with harmonic analysis (Alexandria, 1993)*,

[82] S. Sato, "Weighted weak type $(1,1)$ estimates for oscillatory singular integrals," *Studia Math.*, vol. 141, no. 1, pp. 1–24, 2000.

[83] E. T. Sawyer, "A characterization of a two-weight norm inequality for maximal operators," *Studia Math.*, vol. 75, no. 1, pp. 1–11, 1982.

[84] ——, "A characterization of two weight norm inequalities for fractional and Poisson integrals," *Trans. Amer. Math. Soc.*, vol. 308, no. 2, pp. 533–545, 1988.

[85] E. M. Stein and S. Wainger, "Oscillatory integrals related to Carleson's theorem," *Math. Res. Lett.*, vol. 8, no. 5-6, pp. 789–800, 2001.

[86] S. Treil and A. Volberg, "Entropy conditions in two weight inequalities for singular integral operators," *Adv. Math.*, vol. 301, pp. 499–548, 2016.

[87] R. Urban and J. Zienkiewicz, "Weak type $(1,1)$ estimates for a class of discrete rough maximal functions," *Math. Res. Lett.*, vol. 14, no. 2, pp. 227–237, 2007.

[88] D. Zheng, "The distribution function inequality and products of Toeplitz operators and Hankel operators," *J. Funct. Anal.*, vol. 138, no. 2, pp. 477–501, 1996.